Robust Wavelet Support Vector Machine for Regression Estimation

Xiao-guang Zhang[1,2], Ding Gao[1], Xing-gang Zhang[2], Shi-jin Ren[3]

[1]College of Mechanical and Electrical Engineering, China University of Mining and Technology, Xuzhou, 221008, P.R China

[2]Department of Electronic Science and Engineering, Nanjing University, Nanjing 210093, P.R China

[3]Zhejiang University National Lab of Industrial Control Technology, Hangzhou, P.R.China

doctorzxg@163.com

Abstract

As we know, M-estimation as objective function can be used to tackle this problem that the performance of wavelet network (WN) is affected by gross error severely, but its influence function is determined by the absolute value of residual, so a key problem is how to choose initial parameters. In this paper combining robust estimation with wavelet support vector machine (WSVM), a robust wavelet support vector regression (WSVR) model is developed. Firstly, a new type of wavelet support vector machine is proved and used to determine appropriate WN structure and initial parameters. It can ensure that there should be bigger absolute value of residual for sample with gross error than that for believable sample; Secondly, M-estimation is used as cost function and a method used to determine the threshold adaptively is put forward, then gradient descent is adopted to tune WN parameters. Simulations illustrate that the regression model not only has the multiscale approximation, but also better robustness and generalization, but for some special case, WSVR in single scale can not determine appropriate WN initial parameters and robust learning is very slow, thus multiscale WSVR is worth researching.

Keyword: support vector regression, robust estimation, wavelet support vector machine, wavelet admissible support vector kernel

I. Introduction

It is well known that regression function is to find a function to formulate the relationship between input-output patterns. Regression function is applied to many fields, such as image recovery, pattern recognition, control system and system identification [1]. In most of applications, expectation function is highly nonlinear and difficult to be mathematically formulated accurately. WN is a common regression method. It is based on wavelet decomposition and has more freedom than wavelet does, so it can obtain better approximation precision. It is non-parameter estimation and can obtain regression function through learning from training samples. But during training regression function, if wrong or imprecise training data is adopted, the learning map will surge acutely and the gross error caused by the wrong mode will largen [2]. So how to decrease the effect of gross error is still an interesting problem and there are several successes [3], [4].

Practically, training data are distributed unevenly and most signals are stacked up by the ones with different frequency, which are very suitable for multiscale learning. So people begin to study multiscale support vector regression (SVR), which is much better than single scale SVR in approximation precision mentioned [5], [6]. SVR has also been shown to have excellent performance for insensitive cost function, and a general rule for choosing cost function for SVR has also been proposed according to error distribution [7],[8], [9]. However, the parameters in those approaches are not properly chosen, the final results may be affected by its parameters. This property has also been mentioned in [10]. The choice of parameters of SVR is not straightforward. In fact, for different samples, the optimal sets of parameters are also different. The method of multiscale approximation mentioned above can be regarded as multiscale SVR established by the linear combination of many kernel functions with different parameters. Determining the parameters of kernel functions is based on experiment or cut and try method, which leads to difficult optimization and complex calculation.

Huber robust function has better performance for some noise distribution. When there are outliers, the robust property of SVR is not distinct [11]. In order to improve the robust property of SVR, reference [12] discussed cost function and the corresponding noise density model based on maximum likelihood method, however, it is very difficult to choose suitable cost function. In reference [11], robust SVR can decrease the effect of outliers, which used robust cost function to modify the weight coefficients in order to decrease the effects of outliers.

In reference [2], [3], the objective function based on M-estimation is put forward to decrease the effects that gross error does on regression model. Its principle is to judge whether the point is normal one or the one with gross error according to the absolute value of residual error. The influence function can be chosen based on the gross error to decrease the effects of outlier case. So the initial parameters determine the final approach precision directly. WN is composed of several wavelet bases with different scale, but it is difficult to choose the network structure and initial parameters. In this paper, WSVM is used to determine the structure and initial parameters of WN which are trained with robust objective function. And the method of modifying learning speed online to quicken the training speed because support vector machine (SVM) is based on statistics learning theory and structure risk minimum. So it can solve the problem how to choose initial parameter of WN mentioned above and has excellent generalization to the unknown data [13]. It can adjust the smoothness of approach function according to punishment coefficient because the residual error is bigger between the values of smooth approach function and sample points with gross error. Since the smoothness of initial values is good, the detail of approach function can not fully embody. In this way, using M-estimation as objective function to train it, WN can possess multiscale and robustness and can approach expected real values better. So this WN is called SVM regression model of robust multiwavelet.

In chapter 2, the theory of robust regression approach and the principle of SVR are introduced. In chapter 3, it is proved that radius basis wavelet is a admissible support vector kernel. In chapter 4, the SVR model of robust multiwavelet is introduced. In chapter 5, using the model proposed in this paper, the regression function is simulated and analyzed. And the results are in chapter 6.

II. M-estimation and SVR

A．M-estimation [1], [2]

To N regression samples $\{(x(i), y(i))\}_{i=1}^{n} \subset R^d \times R$, use wavelet neural network (WNN) with single output $f(x)$ as regression approach. Its parameter set is θ, which is adjusted through minimizing objective function $\sum_{i=1}^{l} E(r_i)$. That is

$$\theta_{k+1} = \theta_k - \eta \sum_{i=1}^{N} \frac{\partial E(r_i)}{\partial \theta} \tag{1}$$

where η is learning ratio, $r_i = f(x_i) - y_i$ is residual error and $E(r_i)$ is expectation value of residual error, and then the gradient $\sum_{i=}^{N} \frac{\partial E(r_i)}{\partial \theta}$ can be written:

$$\sum_{i=1}^{N} \frac{\partial E(r_i)}{\partial \theta} = \sum_{i=1}^{N} \frac{\partial E(r_i)}{\partial r_i} \frac{\partial r_i}{\partial \theta} = \sum_{i=1}^{N} \psi(r_i) \frac{\partial f(x_i)}{\partial \theta} \tag{2}$$

$\psi(r_i)$ is called influence function which determines the performance of the function. According to LS rule, $\psi(r_i)$ is equal to r_i. When there are outlier cases, the regression location is far from the real location and the residual error of outlier cases location could be very big. At the same time, the influence function will far bigger than 0. Because of the minimum objective function, LS rule can not be used to approach real curve.

In order to decrease the effects of outlier cases, M-estimation can be used as the objective function of the network and its form is as follows:

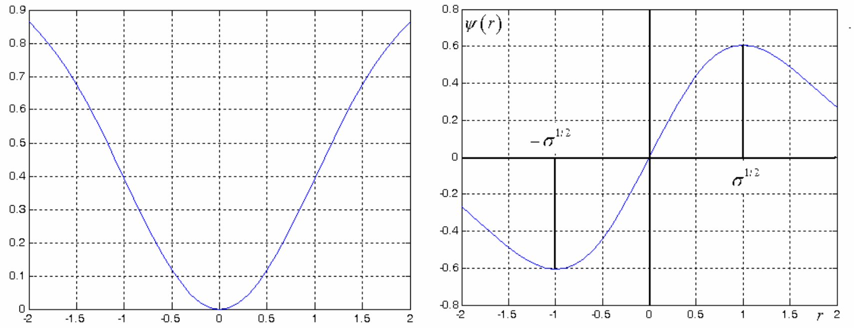
$$\min_{\theta} \sum_{i=1}^{N} \rho(r_i)$$

Derivate it and obtain the minimal value. Suppose $\psi(r_i) = d\rho(r_i) / dr_i$, $\sum_{i=1}^{l} \psi(r_i) = 0$.

There is several M-estimation which can be used as objective function, such as Hampel M-estimation. And the M-estimation like Hampel form can be constructed. In this paper, M-estimation $\rho(r)$ proposed in reference [14] is used as the objective function. Its form and its derivative can be shown in fig.1. The resolution form is

$$\psi(r) = re^{-r^2/2o} \tag{3}$$
$$\rho(r) = \sigma(1 - e^{-r^2/2o})$$

σ is threshold with the change of time, which changes with the residual error according to the training. In fig.1, $\pm\sigma^{1/2}$ are the extremum point of $\psi(r)$, and the turning point of residual error. The objective function $\rho(r)$ and influence function $\psi(r)$ is related the residual error directly. When the residual error exceeds a certain range, the effects on the final parameters decrease step by step. When residual error is in a narrow range, it does approximatively linear effects on the parameters. The initial values of neural network do large effects on the residual error. When the initial values are not good, the residual error in outlier cases will be smaller. But the residual error in accurate points is bigger and can not approach expectation curve well.



(a) **(b)**

Fig.1 (a) $\rho(r)$ and its derivative. (b) $\psi(r)$

B. Support Vector Regression

To N training samples, SVM chooses rational network structure to satisfy the precision condition through the following optimization problem [15]:

$$\min \quad L(w,b) = \frac{1}{2} w^T w + \frac{C}{N} \sum_{i=1}^{N} | y_i - f(x(i)) |_\varepsilon \tag{4}$$

where C is punishment coefficient, which is the compromise of model error and training sample error and can prevent fitting the training data overly to decrease the generation of the model. ε is the given approach precision. n is the length of sample data. $\frac{1}{2} w^T w$ presents model complexity and $|\cdot|_\varepsilon$ is ε non-sensitive function. Through introducing Lagrange factors a_i, a_i^* , the parameters of SVM can be obtained by solving the following dual optimization problem:

$$L(a,a^*) = \frac{1}{2} \sum_{i,j}^{N} (a_i - a_i^*)(a_j - a_j^*) K(x(j), x(i)) - \sum_{i=1}^{N} (a_i - a_i^*) y(i) + \varepsilon \sum_{i=1}^{N} (a_i + a_i^*) \tag{5}$$

Satisfies the restrict conditions

$$\sum_{i=1}^{N} (a_i - a_i^*) = 0$$

$$a_i, a_i^* \in [0, C/N\}$$

In the end, the estimation model of SVR can be obtained as follows:

$$f(x(k)) = \sum_{i=1}^{n_{sv}} (a_i - a_i^*) K(x(k), x(i)) + b \tag{6}$$

where b is warp and the samples corresponding to $a_i - a_i^* \neq 0$ are support vectors, supposing the number of support vectors is n_{SV}. The kernel function $K(x(k), x(i))$ satisfies the Mercer condition and is the of nonlinear map, $\phi : R^d \to F$ which maps $x(k), x(i)$ to the dot metrix of high-dimension feature space $\phi(x(k))$ and $\phi(x(i))$.

$$K(x(k), x(i)) = \phi^T(x(k))\phi(x(i)) \tag{7}$$

where, the nonlinear map $\phi(x)$ is to map sample x to high-dimension feature space and to make it be divided linearly.

III. Wavelet Admissible Support Vector Kernel

Wavelet possesses excellent properties of local approach and multiscale. WN of radius base combines the multiscale property of wavelet and the self-learning of neural network. So it is not only can approach signal in multiscale way, but also can adaptively adjust weights, scale parameters of network and shift parameters according to the data. However, WN has the following shortcomings: (1) It is difficult to determine the initial parameters of initial WN when the input dimension is higher. (2) When there are outliers to samples, it is easy to overfitting. Although references put forward robust learning algorithms of neural network, there is a common problem on how to determine appropriate initial values of parameters [2], [14]. (3) It is difficult to determine the structure of WN.

The kernel function of SVM is admissible support vector kernel. Since it is a good method that the linear learning machine is exceeded to nonlinear learning machine, it should be paid more attention to in machine learning. If wavelet is admissible support vector kernel, so WSVR with wavelet admissible kernel can inherit the merits of wavelet function and may improve the approach precision. WN approaches the signal in multiscale. Since WSVR has similar structure of the WN, use WSVR to determine the structure and initial parameters of WN and the robust loss function as cost function and adopt the method of gradient decrease to adjust the parameters in WN. In this way, this model possesses better generalization, multiscale and robustness properties, and improves the training speed. This model is called robust multiscale SVR, which makes full use of both merits and overcomes their shortcomings.

The following theorems give the form of wavelet admissible support vector kernel.

Theorem 1[1], [3]: Suppose $\psi(t)$ is mother wavelet, a, b_i are translation parameters of scale parameter kernel respectively, and $x, z \in R^d$. The admissible support vector kernel is

$$K(x,z) = \prod_{i=1}^{d} \psi\left(\frac{x_i - b_i}{a}\right)\psi\left(\frac{z_i - b_i}{a}\right) \tag{8}$$

and the wavelet translation invariant kernel

$$K(x,z) = \prod_{i=1}^{d} \psi\left(\frac{x_i - z_i}{a}\right) \tag{9}$$

is also admissible support vector kernel. The formula (8) above is wavelet support vector kernel.

From the theorem 1, one-dimension admissible wavelet support vector kernel is $K(x,z) = \psi(\frac{x-z}{a})$ and it is also the translation invariant kernel. Radius basis wavelet is also admissible support vector kernel, which will be discussed in the following theorem.

Theorem 2: Suppose mother wavelet is $K(x_i, z_i) = \{1-(\frac{x_i - z_i}{a})^2\}\exp\{-\frac{(x_i - z_i)^2}{2a^2}\}$, where $x, z \in R^d$, then the following formula is also wavelet admissible support vector kernel.

$$K(x,z) = \{d - \sum_{i=1}^{d}(\frac{x_i - z_i}{a})^2\}\exp\{-\frac{1}{2}\sum_{i=1}^{d}(\frac{x_i - z_i}{a})^2\} \quad (10)$$

This kernel function is similar to RBF kernel function. $K(x,z) = K(\| x - z \|_2, 0)$, it is means that the function value is related to Euclid distances of input variants. The kernel function can be called radius basis wavelet support vector kernel.

Proof: from the formula $K(x,z) = K(\| x - z \|_2, 0)$, it satisfies the translation invariant.

Based on the reference [2], in order to prove that equation (9) is admissible support vector kernel, it is necessary to prove the Fourier transform to be positive.

From formula (9),

$$K(x,z) = \{d - \sum_{i=1}^{d}(\frac{x_i - z_i}{a})^2\}\exp\{-\frac{1}{2}\sum_{i=1}^{d}(\frac{x_i - z_i}{a})^2\}$$

Its Fourier transform is

$$\begin{aligned}
F[K](\omega) &= (\frac{1}{2\pi})^{-\frac{d}{2}}\int_{R^d}\exp\{-j(<\omega, x>)\}K(x)dx \\
&= (\frac{1}{2\pi})^{-\frac{d}{2}}\int_{R^d}\exp\{-j(<\omega, x>)\}(d - \sum_{i=1}^{d}(\frac{x_i}{a})^2)\exp\{-\frac{1}{2}\sum_{i=1}^{d}(\frac{x_i}{a})^2\}dx_1 \cdots dx_d \\
&= (\frac{1}{2\pi})^{-\frac{d}{2}}[d\prod_{I=1}^{d}\int_R \exp(-j\omega_i x_i)\exp(-\frac{1}{2}(\frac{x_i}{a})^2)]dx_i - \\
&\sum_{i=1}^{d}\int_{R^d}\exp(-j<\omega\cdot x>)(\frac{x_i}{a})^2\exp\{-\frac{1}{2}\sum_{i=1}^{d}(\frac{x_i}{a})^2\}dx_1\cdots dx_d \\
&= d(2\pi)^{\frac{d}{2}}a^d e^{-\frac{1}{2}\sum_{i=1}^{d}\omega_i^2} - \sum_{i=1}^{d}(2\pi)^{\frac{d}{2}}a^d e^{-\frac{1}{2}\sum_{i=1}^{d}\omega_i^2}(1-\omega_i^2) \\
&= (2\pi)^{\frac{d}{2}}a^d e^{-\frac{1}{2}\sum_{i=1}^{d}\omega_i^2}\sum_{i=1}^{n}\omega_i^2
\end{aligned}$$

If $a > 0$, $F[K](\omega) \geq 0$. So K is admissible support vector kernel.

End.

IV. SVR Based on Robust Wavelet

Neural network adopts M-estimation to eliminate the effects of outliers in common cost function. It uses estimation error to determine the value of influence function and separates outliers from most points by decreasing the effects of those points with big estimation error. So it is a key problem to determine appropriate initial parameters of neural network. Although the initial weights can be determined with the routine training methods and then are trained with robust learning algorithm, it is difficult to determine what time is the appropriate transforming time. Since training samples possess multiscale learning and the case of outliers, WN with multiscale is combined with SVR, and using robust estimation as cost function, then the robust multiscale network of WSVR can be constructed.

Suppose there are N samples, which learns with wavelet SVR and obtain n_{sv} support vectors. The corresponding support vector is b_k , and w_k is corresponding weight coefficients of the support vector, where $k = 1,2,\cdots,n_{sv}$. Adopt $a_1 = a_2 = \cdots = a_{n_{sv}} = a$ as scale parameters. Then the WN corresponding to wavelet SVR can be written as follows:

$$f(x_i) = \sum_{k=1}^{n_{sv}} w_k K_k(x_i) + b \tag{11}$$

where $K_k(x_i) = \varphi(\frac{x_i - b_k}{a_k}) = \{d - \frac{\| x_i - b_k \|^2}{a_k^2}\}\exp\{-\frac{\| x_k - b_k \|^2}{2a_k^2}\}$

The cost function can be obtained using the robust objective function mentioned in chapter 2. Then according to cost function, the scale parameters, shift parameters and weights are modified. From the equation (1), there are

$$w_k(t+1) = w_k(k) + \eta\Delta w_k(t) \tag{12}$$

$$a_k(t+1) = a_k(t) + \eta\Delta a_k(t) \tag{13}$$

$$b_k(t+1) = b_k(t) + \eta\Delta b_k(t) \tag{14}$$

From the equation (2), there are

$$\Delta w_k(t) = \frac{\partial E}{\partial w_k} = \frac{1}{N}\sum_{j=1}^{n_{SV}} \psi(r_j(t))[-k_k(x_j)]$$

$$\Delta b_k(t) = \frac{\partial E}{\partial b_k} = \frac{1}{N}\sum_{j=1}^{n_{SV}} \psi(r_j(t))\frac{\partial f(x_j)}{\partial b_k}$$

$$= -\frac{1}{N}\sum_{j=1}^{n_{SV}} \psi(r_j(t)) w_k(\frac{x_j - b_k}{a_k^2})[-2 + d - \frac{\| x_j - b_k \|^2}{a_k^2}]\exp(-\frac{\| x_j - b_k \|^2}{2a_k^2})$$

$$\Delta a_k(t) = \frac{\partial E}{\partial a_k} = \frac{1}{N}\sum_{j=1}^{n_{SV}} \psi(r_j(t))\frac{\partial f(x_j)}{\partial a_k}$$

$$= -\frac{1}{N}\sum_{j=1}^{n_{SV}} \psi(r_j(t)) w_k(\frac{x_j - b_k}{a_k^3})[-2 + d - \frac{\| x_j - b_k \|^2}{a_k^2}]\exp(-\frac{\| x_j - b_k \|^2}{2a_k^2})$$

where $a(t)$ and $b(t)$ depend on the median value of the time. In this paper, the following method can be used to determine σ. Firstly, the percent q of outliers in samples should be determined. Then the value can be estimated as following:

(1) Compute residuals: $r_i(t) = f(x_i) - y_i$, $i = 1,\cdots,N$.

(2) Sort $| r_j(t) |$ by ascending: $| r(t) |_{(1)},\cdots,| r(t) |_{(1-q)N},\cdots,| r(t) |_{(N)}$, where $\sigma^{1/2} = | r(t) |_{(1-q)N}$, q is outlier percent in samples. Because the method of gradient descent converges slowly, the η method which is put forward in reference [4] and to determine optimization learning rate is adopted in this

paper, in order to increase the speed. It determines η according to the residuals of the next step every time.

V. Simulation Analysis and Results

Example1. In order to validate the algorithm proposed in this paper, the sine function with gross error is considered:

$$y = \frac{\sin(x)}{x} + e(x), \quad x \in [-2,2]$$

where $e(x)$ is gross error whose amplitudes is 0.8. The sample number of gross error is 10% of total samples. Choose 50 points as learning samples, including 5 gross error points. And another 50 points is chose as testing samples. Adopt wavelet SVR to determine initial parameters of WN, whose estimation output is shown in fig.2 and the method is detailed in reference [11]. —— denotes the forecast output with certain initial values and —— denotes sample data. Use the algorithm of robust learning introduced in chapter 2 and chapter 5 to learn network parameters. The final output of forecast results is shown in fig.3. Here, —— denotes expectation curve , ——denotes the curve of regression function obtained by learning and —.—denotes the curve of the samples with gross error. From fig.2, wavelet SVR can determine the initial parameters of network well. From fig.3 algorithm, multiscale wavelet network can be obtained by robust learning and can reduce the effects of outliers effectively. And it can approximate the original system well. It can be found that the proposed RSVR can reduce the overfitting phenomena. These results consist with the concept discussed in [11].

From the regression curve shown in fig.3, it can fit sample points and function y well. Under the condition with noise, the simulation results show that robust wavelet support vector machine possesses perfect generalization.

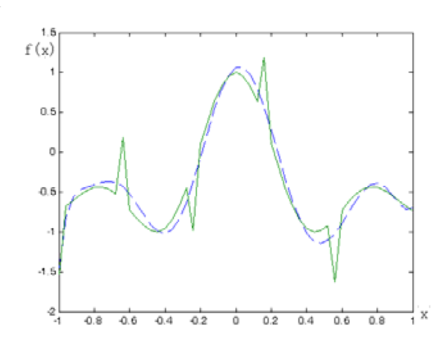


Fig. 2 Forecast output using WSVR to determine initial parameters

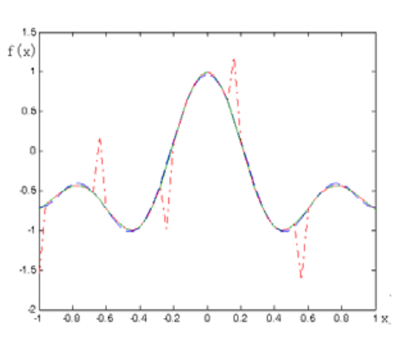


Fig.3 Approach results using WSVR based on robust multiscale

Example2. Take the nonlinear system [16] described by the following difference equation into account.

$$y(k) = 0.3y(k-1)0.6y(k-2) + f[u(k)] + e(t)$$

where $e(t)$ is white noise when the mean value is zero and variance is 0.2. $f(u(k)) = u^3(k) + 0.3u^2(k) - 0.4u(k)$, $u(k) = 0.3\sin(2k\pi/95) + 0.75\sin(2k\pi/175)$. 6% samples outputs are added up to gross errors. There are 140 samples. The prediction output trained by the SVR model of robust WN is shown in fig.4. From fig.4, the prediction is deviated greatly from the system output because of the effect of outliers. The prediction using the algorithm proposed in this paper is shown in fig.5. The result shows that this algorithm can overcome the influence of outliers and approach the system better.

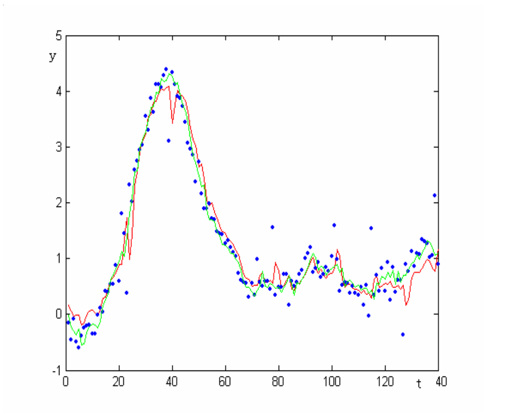


Fig.4 SVR approximation result of wavelet

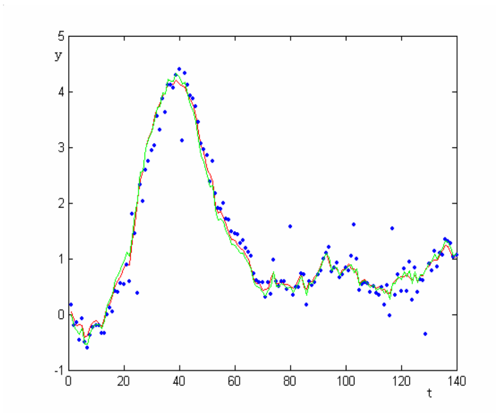


Fig.5 SVR approximation result of robust multiwavelets

Example3. Our experiment is with the chaotic laser time-series from the Santa Fe time-series competition [17]. This is a particularly difficult time-series to predict, due to its chaotic dynamics and the fact that only three "intensity collapse" The training data consists of 70 samples, with the test data being the subsequent 30 samples. The task is to predict the time series by the proposed algorithm, and to see if the proposed algorithm is effective to difficult and complex system. In fig.6, the sum of abstract value of residuals is 0.4488, we can conclude that WSVR in single scale can't satisfactorily approximate system. Based on initial parameters learned from WSVR, the approximation precision becomes higher by robust learning, and the sum of abstract value of residuals is 0.2362. But the some large residual can't reduce, because the large residuals are regard as gross errors. Thus the proposed algorithm can't always obtain initial proper parameters for WN, it is necessary to find multiscale WSVR algorithm.

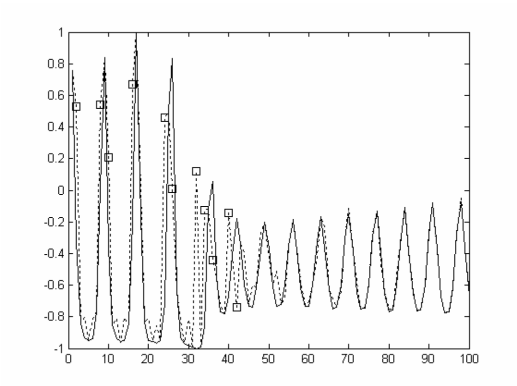


Fig.6 training result using SVM

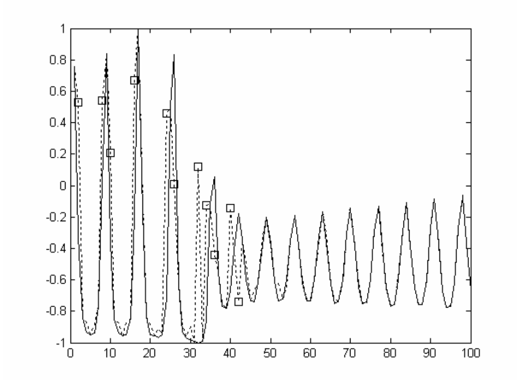


Fig.7 learning result using robust

VI. Conclusion and Future Work

Since there are outliers in samples, SVR is suitable for single scale and has not good robustness when there are outliers. According to the multiscale and multiprecision property of practical samples, the robust WSVR estimation method of regression function is proposed using WN with multiscale property, neural network learning and general approach. And wavelet SVR is used to determine the structure and initial parameters of WN, which makes the samples with outlier cases have big residual error. Then an M-estimation is used as objective function and a method is put forward to determine the thresholds adaptively. The WN with multiscale property is combined with SVR and the robust estimation is regarded as cost function. In this way, the network of robust multiscale WSVR comes into being. And it can adjust the learning speed to increase the training speed online. This model possesses not only multiscale approach, but excellent robustness and better generalization when there are outliers. Simulation results show that this algorithm also possesses highly theoretical and practical value to the research of SVM.

Although the presented algorithm can effectively resolve initial parameters of wavelet network in robust regression, it is noting but that WSVR is approximate signal on single scale, thus it leads to two problems. First it can’t effectively approximate multiscale signal and may result in large training error in believable samples, even have larger error in normal sample than that in outlier. Believable sample can be considered as outlier mistakenly; Secondly, since training errors are large in normal samples when approximating multiscale signal, wavelet network converge slowly and need long time in robust learning. In summary algorithm on WSVR approximating signal in multiscale is worth researching in future.

Acknowledgment

I would like to thank to Dr. Jia-jun Lin and Xing-gang Zhang, they give us lots of advices on theory and technology field, and provide good experimental conditions. The master Yu Li had also helped to finish so many works in the simulation and algorithm study. Now I express the sincerest thanks for them help. This work was partly supported by Jiangsu Planned Projects for Postdoctoral Research.

References

[1] R.-C. Chen, P.-T. Yu. “Fuzzy selection filters for image restoration with neural learning”, *IEEE Trans. Signal Processing*, vol. 47, (1999),pp. 1445–145.

[2] Hung-Hsu Tsai, Pao-Ta Yu. “On the optimal design of fuzzy neural networks with robust learning for function approximation”, *IEEE Trans on. System Man and Cybernetics-Part B: Cybernetics*, vol.30(1),pp.217-233.

[3] H.-H. Tsai. *Design and Analysis of Neuro-Fuzzy Systems for Color Image Restoration and Function Approximation*. Ph.D. dissertation, Dept. Computer Science and Information Engineering, Chung Cheng Univ, Taiwan, R.O.C., Apr. (1999).

[4] David S. Chen, Ramesh C. Jain. “A robust back propagation learning algorithm for function approximation”, *IEEE Trans on. Neural Network,* (1994), vol.3, pp. 467-479.

[5] Michaël A van Wyk, Tariq S. Durrani. “A Framework for Multiscale and Hybrid RKHS-Based Approximators”, *IEEE Trans on Signal Processing*, vol.48(12), pp.3559-3568.

[6] Xuhui Shao, Vladimir Cherkassky. “Multi-Resolution Support Vector machine”,. *1999 International Joint Conference on Neural Networks*, vol.2, pp.1065 – 1070.

[7] A. J. Smola, B. Schölkopf. *A tutorial on support vector regression*. Royal Holloway College, London, U.K., Neuro COLT Tech. Rep. TR-1998-030, 1998.

[8] A. J. Smola, B. Schölkopf, and K. R. Müller. “General cost functions for support vector regression”, presented at the ACNN, *Australian Congr. Neural Networks*, 1998.

[9] A. J. Smola. *Regression estimation with support vector learning machines*. Master's thesis, Technical Univ. Munchen, Munich, Germany, 1998.

[10] J. A. K. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle. *Weighted least squares support vector machines: Robustness and sparse approximation*. Neurocomput., 2001, to be published.

[11] Chen-Chia Chuang,etc, "Robust Support Vector Regression Networks for Function Approximation with Outliers", *IEEE Trans on Neural Networks*, vol.13(6), pp.1322-1329.

[12] Alex J. Smola. *Learning with Kernels*. Thesis, 1998.

[13] Vapnik V. *The nature of statistical learning theory*. SPRINGER press, New York, 1995.

[14] Chien-Cheng Lee etc. Robust radial basis function neural networks, *IEEE Trans on Systems, Man and Cybenetics*, vol.20,1999 (6), pp.624-634.

[15] Du Shu-xin, Wu Tie-ju. "Support Vector Machines for Regression", *Journal of System Simulation (Chinese)*, vol. 15, (2003), pp.1580-1585.

[16] Lv Li-hua. *The research of model method based on wavelet network and robust estimation in complex industrial system*. Doctorial thesis, Zhe Jiang Universtity, 2001.

[17] A. S. Weigend and N. A. Gershenfeld, Eds., *Time Series Predicion:Forecasting the Future and Understanding the Past. Reading*. MA: Addison-Wesley, 1994.

Xiao-guang Zhang born in 1963, received the B.S. degree from the Department of Automation, China University of Mining and Technology, Xuzhou, China, in 1986, and Ph.D. degrees in information engineering from the East China University of Science and Technology, in 2003. Since August 2003, he was engage on postdoctoral research at Institute of Applied Physics Nanjing university. And since August 1998, he has been an Assistant Professor at the College of Mechanical and Electrical Engineering of China University of Mining and Technology. His current research interests include image processing, pattern recognition, process control and intelligent instrument. Dr. Zhang is a Member of Jiangsu of the Meter and Instrument Acad.

Ding Gao born in 1965, received the B.S. degree from Shanghai Jiaotong university in 1986, and Ph.D. degrees in China University of Mining and Technology, in 2005. And since 1999, he has been an Assistant Professor at the College of Mechanical and Electrical Engineering of China University of Mining and Technology. His current research interests include intelligence control, design of Mechanical and Electrical system, weld and manufacture.