ATLANTIS PRESS
Advances in Engineering Research (AER), volume 148
3rd Workshop on Advanced Research and Technology in Industry Applications (WARTIA 2017)

Modeling Based on Smooth Support Vector Regression with ICA Feature Extraction

Xu-Sheng Gan[1,a], Jun Han [2,b], Zhi-bin Chen [2,c]

[1] XiJing College, Xi'an, Shaanxi, 710123, China
[2] Air Traffic Control and Navigation College, Air Force Engineering University, Xi'an, Shaanxi, 710051, China

[a] ganxusheng123@163.com, [b] jun_han@163.com, [c] zhibin_chen@sina.com

Keywords: Smooth support vector regression; Feature extraction; Independent component analysis; Quadratic optimization;

Abstract. Smooth Support Vector Regression (SSVR) is new modified edition of traditional support vector regression for better performance. To further improve the modeling capability of SSVR, it is necessary to take into account the feature extraction based on Independent Component Analysis (ICA) before SSVR. Simulation on the example of function approximation shows that the result of SSVR based on ICA feature extraction is better than that of SSVR without ICA preprocess.

Introduction

In the Support Vector Regression (SVR), the input dimensions, noise contained in the samples and multiple correlations among the variables have a great influence on the performance and quality of the modeling [1][2]. To reduce the complexity of learning problems and improve the generalization ability of the SVR model, it is necessary to carry on feature extraction for sample data before SVR modeling. In the paper, Smooth Support Vector Regression (SSVR) is thought as our research object, and Independent Component Analysis (ICA) is introduced to realize the feature extraction before SSVR.

ICA algorithm based on joint diagonalization of characteristic matrix

ICA applies the high order statistic information in feature extraction of the data, and emphasizes the independence between data characteristics [1]. Using ICA to process the original training samples can eliminate high order correlation of the data, remove the redundant features, mapped the input pattern space in the corresponding independent component space [2]. At the same time, because of the high order independence among each independent component, the regression model obtained has the very strong generalization ability.

ICA usually assumes that the observation signal is from the linear mixing by several background source, the processing model is

$$x = As \tag{1}$$

where x is M-dimension observation vector, s is N-dimension implicit vector (or source), A is the $M \times N$ linear mixed matrix. Under the assumption of statistical independence, ICA seeks a linear transformation W to make

$$y = \hat{s} = Wx \tag{2}$$

where the observation vector x is known, but W and s are unknown.

At present, There are a variety of ICA algorithms, JADE algorithm based on the joint approximation diagonalization of characteristic matrix, proposed by Cardoso, is an stable, robus algebra ICA method numerically. it first whiten the observation signal, i.e.

$$z(t)=Wx(t)=W[As(t)+n(t)]=Us(t)+Wn(t) \tag{3}$$

Thus, the problem of determining a $M\times N$ matrix A is converted into the problem of determining a $N\times N$ unitary matrix V. W is the whitening matrix. Estimation of Y depends on the high order cumulant (usually 4 order). For any $N\times N$ matrix, 4 order cumulant matrix can be defined as

$$N=Q_z(M)\Leftrightarrow n_{ij}=\sum_{k,l=1,n} Cum(z_i,z_j^{\bullet},z_k,z_l^{\bullet})m_{lk} \tag{4}$$

where $1\le i$, $j\ge n$. By maximizing the following reference function to achieve the joint approximation diagonalization of the cumulant matrix set $\hat{N}^e=\{\hat{\lambda}_r\hat{M}_r \mid 1\le r\le n\}$, thus we can get the unitary matrix.

$$C(V,N)=\sum_{r=1,s}|\,diag(V^H N_r V)\,|^2 \tag{5}$$

Smooth support vector machine for the regression

SSVR is a great improvement of the traditional SVR, it can transformed a constrained quadratic optimization problem into an unconstrained convex quadratic optimization problem in SVR algorithm, so as to effectively reduce the training complexity of support vector machine, improve the training speed, decrease the internal memory used. SSVR algorithm not only inherits the excellent generalization ability of SVR algorithm, but also compared with the SVR algorithm has better fitting speed and prediction accuracy. It is very suitable for the modeling problem complicated nonlinear dynamic system [5][6].

Suppose that $X\in R^n$ is the input space, $Y\in R$ is the output space. Given the training sample set $\{(x_1,y_1),\cdots,(x_l,y_l)\}\in X\times Y$, The purpose of SVM regression is to adopt the function $f(x)=w\cdot\phi(x)$ $+b$ to fit the sample set, also require the good generalization ability of SVM. $\phi(x)$ is a nonlinear mapping from the input space to the feature space. The paper discusses the support vector regression under the quadratic ε insensitive loss function. Through the changing of ε value, we can control the sparsity of the solution.

SVM regression of quadratic ε insensitive loss function can be expressed as the following quadratic optimization problems with constraints

$$\min\ \frac{1}{2}w^T w+C\sum_{i=1}^{l}\xi_i^2+C\sum_{i=1}^{l}(\xi_i^*)^2 \tag{6}$$

$$\begin{aligned} s.t.\ \ & w^T\phi(x_i)+b-y_i\le\varepsilon+\xi_i \\ & y_i-w^T\phi(x_i)-b\le\varepsilon+\xi_i^* \\ & \xi_i,\xi_i^*\ge 0, i=1,\cdots,l \end{aligned}$$

where $C>0$ isthe tradeoff parameter between the function complexity and the loss error.

In the smoothing method, the canonical factor C is transformed into $C/2$, and $b^2/2$ is added, which has almost no influence on the solution of the original problem, i.e.

$$\min_{w,b,\xi,\xi^*}\ \frac{1}{2}(w^T w+b^2)+\frac{C}{2}\sum_{i=1}^{l}\xi_i^2+\frac{C}{2}\sum_{i=1}^{l}(\xi_i^*)^2 \tag{7}$$

$$\begin{aligned} s.t.\ \ & w^T\phi(x_i)+b-y_i\le\varepsilon+\xi_i \\ & y_i-w^T\phi(x_i)-b\le\varepsilon+\xi_i^* \\ & \xi_i,\xi_i^*\ge 0, i=1,\cdots,l \end{aligned}$$

If $z_i=w^T\phi(x_i)+b-y_i-\varepsilon$, $z_i^*=y_i-w^T\phi(x_i)-b-\varepsilon$, and let $\xi_i=(z_i)_+$, $\xi_i^*=(z_i^*)_+$, where $(u)_+=$ $\max|u,0|$, then the formula (7) can be transformed as the following quadratic optimization problems:

$$\min_{w,b} \frac{1}{2}(w^T w + b^2) + \frac{C}{2}\sum_{i=1}^{l}(z_i)_+^2 + \frac{C}{2}\sum_{i=1}^{l}(z_i^*)_+^2 \tag{8}$$

It is easy to see that the formula (8) is an unconstrained convex quadratic optimization problem, having the only solution.

However, the objective function in the formula (8) is not two times differentiable. Therefore, we use a strictly convex and infinitely differentiable smooth function

$$p(u,\alpha) = \frac{1}{\alpha}\ln(1+e^{\alpha u}), \alpha > 0 \tag{9}$$

instead of $(u)_+$.

The studies show that, when the smoothing parameter $\alpha = 10$, $p(u,\alpha)$ curve almost coincides with $(u)_+$. And when $\alpha \to \infty$, $p(u,\alpha)$ can arbitrarily approximate $(u)_+$. Then we get the smooth SVM regression

$$\min_{w,b} \frac{1}{2}(w^T w + b^2) + \frac{C}{2}\sum_{i=1}^{l} p(z_i,\alpha)^2 + \frac{C}{2}\sum_{i=1}^{l} p(z_i^*,\alpha)^2 \tag{10}$$

Let $\tilde{w} = (w,b)$, the formula (3) and (4) type can be written as the following norm form

$$\min_{\tilde{w}\in R^{n+1}} \Phi(\tilde{w}) := \frac{1}{2}\|\tilde{w}\|^2 + \frac{C}{2}\|(z)_+\|^2 + \frac{C}{2}\|(z^*)_+\|^2 \tag{11}$$

$$\min_{\tilde{w}\in R^{n+1}} \Phi_\alpha(\tilde{w}) := \frac{1}{2}\|\tilde{w}\|^2 + \frac{C}{2}\|p(z,\alpha)\|^2 + \frac{C}{2}\|p(z^*,\alpha)\|^2 \tag{12}$$

It can be validated that, when $\alpha \to \infty$, the solution of the formula (12) converge to the solution of the formula (11).

We give the Quasi-Newton-Armijo algorithm of Smooth support vector machine for the regression as follows:

1. Any initialization iteration points $\tilde{w}^1 \in R^{n+1}$, the calculation accuracy setting $\varepsilon > 0$;
2. Construct Quasi-Newton iteration direction: $d^i = -H_i \nabla\Phi_\alpha(\tilde{w}^i)$, $s.t.$ $z^T H_i z \ge \beta\|z\|^2$, $\beta > 0$, $\forall z \in R^{n+1}$;
3. Select the Armijo iteration step: $\tilde{w}^{i+1} = \tilde{w}^i + \lambda_i d^i$, where $\lambda_i \in \{1, 1/2, 1/4, \cdots\}$ is maximum that meets the $\Phi_\alpha(\tilde{w}^{i+1}) - \Phi_\alpha(\tilde{w}^i) \le \delta\lambda_i \nabla\Phi_\alpha(\tilde{w}^i) d^i$, and $\delta \in (0, 1/2)$;
4. If $\|\nabla\Phi_\alpha(\tilde{w}^i)\| \le \varepsilon$, then the iterations is terminated.

Integrated structure SSVR based on ICA feature extraction

ICA-SSVR is actually a hybrid regression algorithm. Firstly, through ICA method, the original samples are processed by ICA to extract the principal features, and then use the obtained principal features to establish the SSVR model to complete the regression. The structure of realization for ICA-SSVR is shown in Fig. 1.

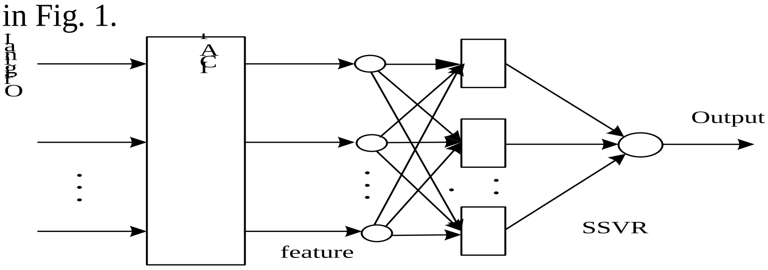

Fig.1 Structure of ICA-SSVR

Numerical example

For a nonlinear piecewise function

$$f(x)=\begin{cases}0.4x+4 & -10\le x\le -2\\ 3.2 & -2\le x<1\\ 3.2e^{-0.5(x-1)}\cos(0.6x^2+0.6x) & 1\le x\le 10\end{cases} \tag{13}$$

We radomly take 481 samples in [-10, 10]. The kernel function can be selected as RBF kernel $K(x_i,x_j)=\exp(-\frac{\|x_i-x_j\|^2}{2\sigma^2})$ and SSVR hyber-parameters: $C=1000$, $\sigma=1.5$, $\varepsilon=2$. After the model is trained, Table 1 shows the comparison of approximation performance between 4 algorithms. Fig.2 and Fig.3 give respectively the approximate effect and error of ICA-SSVR, in Fig.2, "—" denotes the reference curve,"---" denotes the approximation curve.

Table 1 Comparison of approximation result between 4 algorithms

Model	RBFNN	IAFSA-RBFNN	SSVR	ICA-SSVR
Accuracy MSE	0.0629	0.0165	0.0229	0.0085
Training time (s)	224.10	124.65	2.17	1.92

It's not hard to see, compared with SSVR without ICA preprocess, RBFNN (RBF neural network), and IAFA-RBFNN (RBFNN based on artificial fish swarm learning algorithm with an adaptive adjustment), SSVR based on ICA feature extraction has a better modeling performance on the accuracy and training speed, it shows that the SSVR modeling technology based on ICA feature extraction can assuredly decrease the system bias brought by the nonlinear characteristics of the system in modeling. This also provide an effective, intelligentized modeling approach for the nonlinear system.

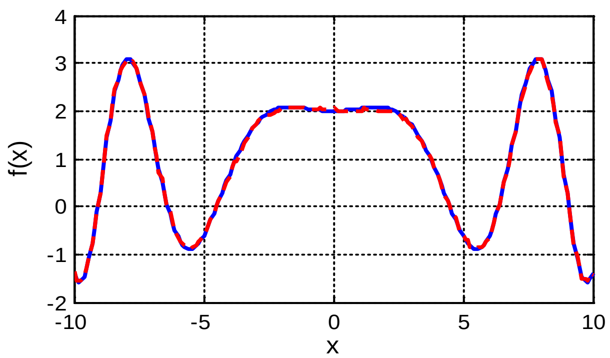

Fig.2 Approximate effect of ICA-SSVR

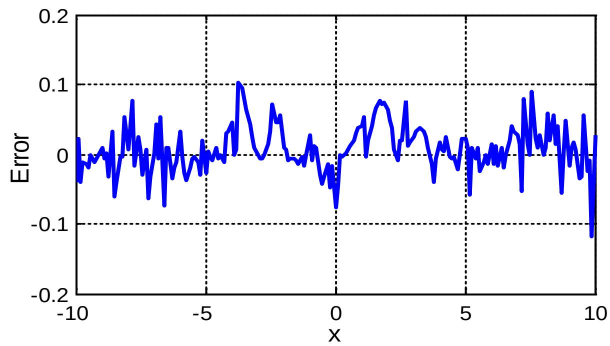

Fig.3 Approximate error of ICA-SSVR

Conclusions

A hybrid modeling method based on SSVR algorithm and ICA feature extraction is proposed. In the method, ICA method is first used to extract the feature from the original samples, and then for these obtained feature carry out the process of SSVR modeling. The example of function approximation indicates that the hybrid modeling method proposed has a good approximation capability.

References

[1] V. Vapnik. The nature of statistical learning theory. Berlin: Springer, (2000)

[2] A. J. Smola, B. Scholkopf. A tutorial on support vector regression. Statistics and Computing, 14(3), (2004), 199-222

[3] W. D. Jiao, S. X. Yang, Z. T. Wu. ICA based neural networks for pattern recognition of mechanical faults. Transactions of the Chinese Society of Agricultural Machinery 35(4), (2004), 151-154

[4] L. Zhong, H. Z. Song. Structural damage identification based on PCA and ICA. Journal of Wuhan University of Technology, 28(7), (2006), 93-96

[5] J. P. Tu, J. Cai. Smooth support vector regression. Journal of Hubei University, 28(1), (2006), 28-31

[6] C. Zhang, P. Han, G. J. Tang. Fault analysis on rotor-to-stator rub-impact for rotary machinery using smooth support vector regression. Chinese Journal of Construction Machinery, 6(1), (2008), 84-89