http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 5 (3): 439-444, 2006
ISSN 1812-5638

Least Squares Support Vector Machine on Morlet Wavelet Kernel Function and its Application to Nonlinear System Identification

Fangfang Wu and Yinliang Zhao
Institute of Neocomputer, Xi'an Jiaotong University, Xi'an 710049, People's Republic of China

Abstract: The kernel function of Support Vector Machine (SVM) is an important factor for the learning result of SVM. Based on the wavelet decomposition and conditions of the support vector kernel function, Morlet wavelet kernel function for SVM is proposed. This function is not only a kind of orthonormal function, but also suitable for local signal analysis, signal-noise separation and detection of jumping signals, thus it enhances the generalization ability of the SVM. According to the wavelet kernel function and the regularization theory, Least squares support vector machine on Morlet wavelet kernel function (LS-MWSVM) is proposed to greatly simplify the solving process of MWSVM. The LS-MWSVM is then applied to the nonlinear system identification to test the validity of the Morlet wavelet kernel function. Computer simulations show that the modeling ability is improved and computation burden is alleviated, comparing with LS-SVM whose kernel function is Gaussian function.

Key words: SVM, support vector kernel function, Morlet wavelet kernel function, LS-MWSVM

INTRODUCTION

Vapnik's support vector machine (SVM) is a statistical model of data that simultaneously minimizes model complexity and data fitting error (Vapnik, 1995). SVM has attracted a lot of interest and has spurred voluminous work in Machine Learning, both theoretical and experimental. The remarkable generalization ability exhibited by SVM can explained through margin-based VC theory (Burges, 1998; Zhang, 2000).

SVM has been a promising method for data classification and regression (Bernhard *et al*., 1997; Edgar *et al*., 1997). Their success in practice is drawn by its solid mathematical foundations which convey the following two salient properties: (1) Margin maximization, the classification boundary functions of SVM maximize the margin, which in machine learning theory, corresponds to maximizing the generalization performance given a set of training data. (2) Nonlinear transformation of the feature space using the kernel trick, SVM handles a nonlinear classification efficiently using the kernel trick which implicitly transforms the input space into another high dimensional feature space.

For pattern recognition and regression analysis, the non-linear ability of SVM can use kernel mapping to achieve. For the kernel mapping, the kernel function must satisfy the condition of Mercer (Mercer, 1909). The Gauss function is a kind of kernel function which is general used. It shows the good generalization ability. However, for our used kernel functions so far, the SVM can not approach any curve in $L_2(R)$ space (quadratic continuous integral space), because the kernel function which is used now is not the complete orthonormal base. This character lead the SVM can not approach every curve in the $L_2(R)$ space, similarly, the regression SVM can not approach every function.

According to the above describing, we need find a new kernel function and this function can build a set of complete base through horizontal floating and flexing. As we know, this kind of function has already existed and it is the wavelet functions. Based on wavelet decomposition, this study propose a kind of allowable support vector's kernel function which is named Morlet wavelet kernel function and we can prove that this kind of kernel function is existent. The Morlet wavelet kernel functions are the orthonormal base of $L_2(R)$ space. At the same time, combining this kernel function with least squares support vector machine (Suykens *et al*., 1999), we can build a new SVM learning algorithm that is Least squares support vector machine on Morlet wavelet kernel function (LS-MWSVM). Finally, the LS-WWSVM is applied to the nonlinear system identification to test the efficiency and validity of the Morlet wavelet kernel.

Corresponding Author: Fangfang Wu, Institute of Neocomputer, Xi'an Jiaotong University, Xi'an 710049, People's Republic of China

SUPPORT VECTOR MACHINE

For the given samples set $\{(x_1, y_1),\ldots,(x_l, y_l)\}$, $x_i \in R^n$, $y_i \in R$, l is the samples number, n is the number of input dimension. In order to approach the function f(x) with respect to this data set precisely, for regression analysis, SVM use the regression function as follows:

$$f(x) = \sum_{i=1}^{l} w_i k(x_i, x) + b \quad (1)$$

w_i is the weight vector and b is the threshold, $k(x_i, x)$ is the kernel function.

Training a SVM can be regarded as to minimize the value of J(w, b):

$$J(w,b) = \min \frac{1}{2}\|w\|^2 + \gamma \sum_{i=1}^{l}\left(y_k - \sum_{i=1}^{l} w_i k(x_i, x) - b\right)^2 \quad (2)$$

The kernel function $k(x_i, x)$ must be satisfied with the condition of Mercer (1909). When we define the kernel function $k(x_i, x)$, we also define the mapping from input to character's space. The general used kernel function of SVM is Gauss function, defined as follows:

$$k(x, x') = \exp(-\|x - x'\|^2 / 2\sigma^2) \quad (3)$$

For this equation, σ is a parameter which can be adjusted by users.

SUPPORT VECTOR'S KERNEL FUNCTION

The conditions of support vector's kernel function: The support vector's kernel function can be described as not only the product of point, such as $k(x,x') = k(<x \cdot x'>)$, but also the horizontal floating function, such as $k(x,x') = k(x-x')$ (Burges, 1999). In fact, if a function satisfied the condition of Mercer, it is the allowable support vector's kernel function.

Theorem 1 (Mercer, 1909): The symmetry function k(x, x') is the kernel function of SVM if and only if: for all function g≠0 which satisfied the condition of $\int_{R^d} g^2(\xi)d\xi < \infty$ we need satisfy the condition as follows:

$$\iint_{R^d \otimes R^d} k(x, x')g(x)g(x')dxdx' \geq 0 \quad (4)$$

This theorem proposed a simple method to build the kernel function.

For the horizontal floating function, because hardly dividing this function into two same functions, we can give the condition of horizontal floating kernel function.

Theorem 2 (Smola *et al.*, 1998; Burges, 1999): The horizontal floating function is a allowable support vector's kernel function if and only if the Fourier transform of k(x) need satisfy the condition as follows:

$$F[k(w)] = (2\pi)^{-\frac{d}{2}} \int_{R^d} \exp(-jwx)k(x)dx \geq 0 \quad (5)$$

Morlet wavelet kernel function: If the wavelet function Ψ(x) satisfied the conditions: $\psi(x) \in L_2(R) \cap L_1(R)$ and $\hat{\psi}(0) = 0$, $\hat{\psi}$ is the Fourier transform of function Ψ(x). The wavelet function group can be defined as:

$$\psi_{a,m}(x) = (a)^{-\frac{1}{2}} \psi\left(\frac{x-m}{a}\right) \quad (6)$$

$m \in R$; $a \geq 0$. For this equation, a is the flexible coefficient, m is the horizontal floating coefficient and Ψ(x) is the base wavelet. For the function f(x), $f(x) \in L_2(R)$, the wavelet transform of f(x) defined as:

$$(W_\psi f)(a,m) = (a)^{-\frac{1}{2}} \int_{-\infty}^{+\infty} f(x)\overline{\psi\left(\frac{x-m}{a}\right)} dx \quad (7)$$

the wavelet inverse transform for f(x) is:

$$f(x) = C_\psi^{-1} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left[(W_\psi f)(a,m)\right] \psi_{a,m}(x) \frac{da}{a^2} dm \quad (8)$$

For the above Eq. 8, C_ι is a constant with respect to Ψ(x). The theory of wavelet decomposition is to approach the function f(x) by the linear combination of wavelet function group.

If the wavelet function of one dimension is Ψ(x), using tensor theory (Zhang *et al.*, 1992), the multidimension wavelet function can be defined as:

$$\psi_d(x) = \prod_{i=1}^{d} \psi(x_i) \quad (9)$$

We can build the horizontal floating kernel function as follows:

$$k(x, x') = \prod_{i=1}^{d} \psi\left(\frac{x_i - x_i'}{a_i}\right) \quad (10)$$

a_i is the flexible coefficient of wavelet, $a_i>0$. So far, because the wavelet kernel function must satisfy the conditions of theorem 2, the number of wavelet kernel function which can be showed by existent functions is few (Zhang *et al.*, 2004). Now, we give an existent wavelet kernel function: Morlet wavelet kernel function and we can prove that this function can satisfy the condition of allowable support vector's kernel function. Morlet wavelet function is defined as follows:

$$\psi(x) = \cos(\omega_0 x)e^{-\frac{x^2}{2}} \tag{11}$$

We define Morlet wavelet kernel function as follows:

$$k(x, x') = \prod_{i=1}^{d} \psi\left(\frac{x_i - x_i'}{a_i}\right) = \prod_{i=1}^{d} \cos\left(\omega_0\left(\frac{x_i - x_i'}{a_i}\right)\right) \exp\left(-\frac{(x_i - x_i')^2}{2a_i^2}\right) \tag{12}$$

Theorem 3: Morlet wavelet kernel function is defined as:

$$k(x, x') = k(x - x') = \prod_{i=1}^{d} \cos\left(\omega_0\left(\frac{x_i - x_i'}{a_i}\right)\right) \exp\left(-\frac{(x_i - x_i')^2}{2a_i^2}\right) \tag{13}$$

and this kernel function is a allowable support vector kernel function.

Proof: According to the theorem 2, we only need to prove

$$F[k(w)] = (2\pi)^{-\frac{d}{2}} \int_{R^d} \exp(-jwx)k(x)dx \geq 0$$

$$F(k(\omega)) = (2\pi)^{-\frac{d}{2}} \int_{R^d} k(x)\exp(-j\omega x)dx$$

$$= (2\pi)^{-\frac{d}{2}} \int_{R^d} \prod_{i=1}^{d} \cos\left(\omega_0 \frac{x_i}{a_i}\right) \exp\left(-\left(\frac{x_i}{a_i}\right)^2\right) \exp(-j\omega x)dx$$

$$= (2\pi)^{-\frac{d}{2}} (a_i)^d \prod_{i=1}^{d} \int_{-\infty}^{+\infty} \cos\left(\omega_0 \frac{x_i}{a_i}\right) \exp\left(-\frac{1}{2}\left(\frac{x_i}{a_i}\right)^2\right) \exp\left(-j\omega a_i \frac{x}{a_i}\right) d\frac{x_i}{a_i}$$

$$= (2\pi)^{-\frac{d}{2}} (a_i)^d \prod_{i=1}^{d} \int_{-\infty}^{+\infty} \cos(\omega_0 t) \exp\left(-\frac{t^2}{2}\right) \exp(-j\omega a_i t)dt$$

$$= (2\pi)^{-\frac{d}{2}} (a_i)^d \prod_{i=1}^{d} \left(\frac{\sqrt{2\pi}}{2}\left(\exp\left(-\frac{(\omega_0 + \omega a_i)^2}{2}\right) + \exp\left(-\frac{(\omega_0 - \omega a_i)^2}{2}\right)\right)\right)$$

$$= \left(\frac{a_i}{2}\right)^d \prod_{i=1}^{d} \left(\exp\left(-\frac{(\omega_0 + \omega a_i)^2}{2}\right) + \exp\left(-\frac{(\omega_0 - \omega a_i)^2}{2}\right)\right)$$

$$F(k(\omega)) \geq 0$$

If we use the support vector' kernel function as Morlet wavelet kernel function, the classifier function of SVM is defined as:

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{l} w_i \prod_{i=1}^{d} \cos\left(\omega_0\left(\frac{x_j - x_j^i}{a_j^i}\right)\right) \exp\left(-\frac{1}{2}\left(\frac{x_j - x_j^i}{a_j^i}\right)^2\right) + b\right) \tag{14}$$

For regression analysis, the output function is defined as:

$$f(x) = \sum_{i=1}^{l} w_i \prod_{i=1}^{d} \cos\left(\omega_0\left(\frac{x_j - x_j^i}{a_j^i}\right)\right) \exp\left(-\frac{1}{2}\left(\frac{x_j - x_j^i}{a_j^i}\right)^2\right) + b \tag{15}$$

x_j^i is the value of ith training sample's jth attribute. Using the Littlewood-Paley wavelet kernel function, we can give the regression function a new concept: using the linear combination of wavelet function group, we can approach any function f(x), that is to say, we can find the wavelet coefficients to decomposition the function f(x). For this study, we give the value 5 to the parameter ω_0 of Morlet wavelet kernel function.

LEAST SQUARES SUPPORT VECTOR MACHINE ON MORLET WAVELET KERNEL FUNCTION

Least squares support vector machine is a new kind of SVM (Suykens *et al*., 1999). It derived from transforming the condition of inequation into the condition of equation. Firstly, we give the linear regression algorithm as follows.

For the given samples set $\{(x_1, y_1),\ldots,(x_l, y_l)\}$, $x_i \in R^n$, $y_i \in R$, l is the samples number, n is the number of input dimension. The linear regression function is defined as:

$$f(x) = w^T x + b \tag{16}$$

Importing the structure risk function, we can transform regression problem into protruding quadratic programming:

$$\min \frac{1}{2}\|w\|^2 + \gamma \frac{1}{2}\sum_{i=1}^{l} \xi_i^2 \tag{17}$$

The limited condition is:

$$y_i = w^T x_i + b + \xi_i \tag{18}$$

we define the Lagrange function as:

$$L = \frac{1}{2}\|w\|^2 + \gamma\frac{1}{2}\sum_{i=1}^{l}\xi_i^2 - \sum_{i=1}^{l}\alpha_i(w^T x_i + b + \xi_i - y_i) \tag{19}$$

According the KKT condition, we can get:

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{l}\alpha_i x_i \tag{20}$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{l}\alpha_i = 0 \tag{21}$$

$$\frac{\partial L}{\partial \xi_i} = 0 \rightarrow \alpha_i = \gamma\xi_i;\quad i = 1,\ldots,l \tag{22}$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow w^T x_i + b + \xi_i - y_i = 0;\quad i = 1,\ldots,l \tag{23}$$

From Eq. 20-23, we can get the following linear equation:

$$\begin{bmatrix} I & 0 & 0 & -x \\ 0 & 0 & 0 & \vec{1}^T \\ 0 & 0 & \gamma I & -I \\ x^T & \vec{1} & I & 0 \end{bmatrix}\begin{bmatrix} w \\ b \\ \xi \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ y \end{bmatrix} \tag{24}$$

$x = [x_1,\ldots, x_l]$, $y = [y_1,\ldots, y_l]$, $\vec{1} = [1, \ldots, 1]$, $\xi = [\xi_1,\ldots, \xi_l]$, $\alpha = [\alpha_1,\ldots, \alpha_l]$.

The equation result is:

$$\begin{bmatrix} 0 & \vec{1}^T \\ \vec{1} & x^T x + \gamma^{-1} I \end{bmatrix}\begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{25}$$

$$w = \sum_{i=1}^{l}\alpha_i x_i,\ \xi_i = \alpha_i/\gamma$$

For non-linear problem, The non-linear regression function is defined as:

$$f(x) = \sum_{i=1}^{l}\alpha_i k(x_i, x) + b \tag{26}$$

The above equation result can be altered as:

$$\begin{bmatrix} 0 & \vec{1}^T \\ \vec{1} & K + \gamma^{-1} I \end{bmatrix}\begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{27}$$

$K = \{k_{ki} = k(x_k, x_i)\}_{k,i=1}^{l}$, the function $k(\cdot)$ is the Morlet wavelet kernel function. Based on Morlet wavelet kernel function, we can get a new learning method: Least squares support vector machine on Morlet wavelet kernel function (LS-MWSVM). In fact, this algorithm is also Least squares support vector machine. We only use the Morlet wavelet kernel function to represent the kernel function of SVM.

There is only one parameter γ need to be made certain for this algorithm and the number of parameters of this kind of SVM is smaller than other kind of SVM, at the same time, the uncertain factors are decreased. Additionally, because using least squares method, the computation speed of this algorithm is more rapid than other SVM.

Because LS-SVM can not optimize the parameters of kernel function, it is hard to select $l\times d$ parameters. For convenience, we fix $a_j^i = a$ and the number of kernel function's parameters is 1. We can use the Cross-Validation method to select the value of parameter a.

APPLICATION OF LS-MWSVM TO NONLINEAR SYSTEM IDENTIFICATION

In this section, to validate the performance of the wavelet kernel, LS-MWSVM is applied to nonlinear system identification. At the same time, the results obtained by the Morlet wavelet kernel are compared with that obtained by the Gaussian kernel.

There are two different examples of nonlinear systems identification, they are all done on an Intel P4 PC (with a 2.0 GHZ CPU and 512MB memory) running Microsoft Windows 2000 Professional, Matlab6.5. At the same time, we use the approaching error as follows (Zhang *et al*., 1992):

$$E = \sqrt{\sum_{i=1}^{l}(y_i - f_i)^2 \Big/ \sum_{i=1}^{l}(y_i - \bar{y}_i)^2},\quad \bar{y} = \frac{1}{l}\sum_{i=1}^{l} y_i \tag{28}$$

Example 1: Consider a unitary nonlinear system.

$$u(t+1) = \frac{1.4e^{u(t)}}{1 + e^{u(t)}} + v(t)^2 \tag{29}$$

$v(t) = 0.4\sin(2\pi t/200) + 0.2\cos(2\pi t/40) + 0.4\sin(2\pi t/1000)$, where $v(t)$ is random input in interval [-1,1].One takes 400 points as the training samples and 80 points as testing samples. The original input $u(1) = 0$.

The result of the identification can be described as Table 1, Fig. 1 and 2. In the Fig. 1 and 2, the real line is the real value of nonlinear system, the solid

Table 1: The unitary nonlinear system identification result

	The parameter of kernel function	Approximation error
Gaussian kernel	$\sigma = 2$	0.0376
Morlet wavelet kernel	$a = 2$	0.0162

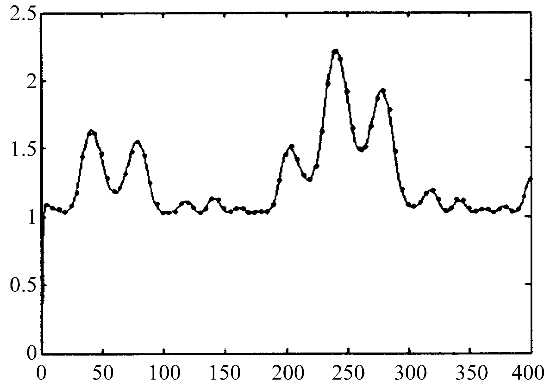

Fig. 1: The unitary nonlinear system identification based on Gaussian kernel

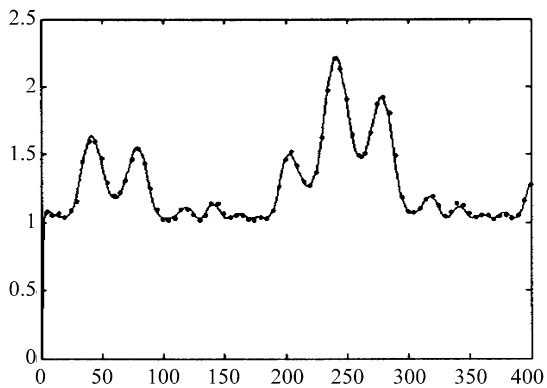

Fig. 2: The unitary nonlinear system identification based on MORLET wavelet kernel

point is the identification's value. Figure 1 is the result of identification which uses the Gaussian kernel function. Figure 2 is the result of identification which uses the Morlet wavelet kernel function. From these result, we can find that the Morlet wavelet kernel function not only has the capacity of non-linear mapping, but also inherit the characters of Morlet wavelet's orthonormal capacity.

Example 2: Consider a twovariable nonlinear system.

$$
\begin{aligned}
u_1(t+1) &= (u_2(t)-u_1(t))/(1+(u_2(t)-u_1(t))^2)+7v(t) \\
u_2(t+1) &= u_1(t)^2/(1+(u_2(t)-u_1(t))^2)-15v(t)
\end{aligned} \tag{30}
$$

$v(t) = 0.8\cos(2\pi t/100) + 0.2\sin(2\pi t/20)$, where $v(t)$ is the random input in interval [-1,1].One takes 400 points as the training samples and 80 points as testing samples. The original input $u_1(1) = 0$, $u_2(1) = 0$.

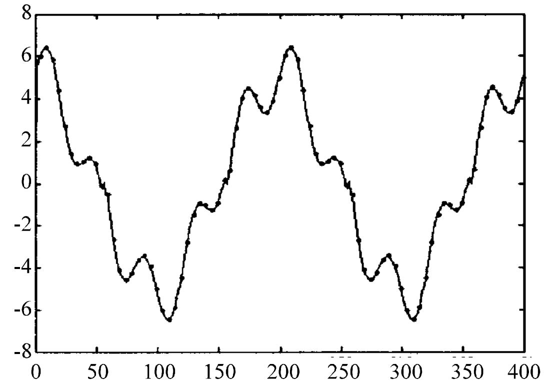

Fig. 3: The twovariable nonlinear system identification of u_1 based on Morlet identification of u_2 based on Morlet

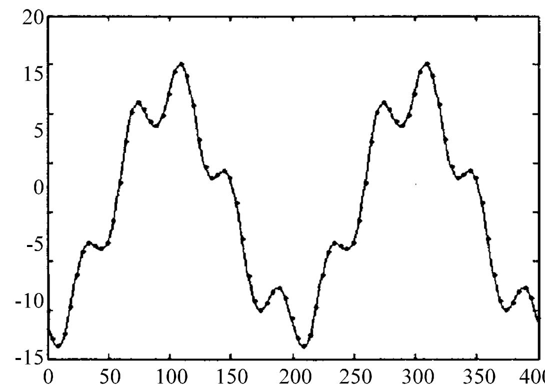

Fig. 4: The twovariable nonlinear system wavelet kernel wavelet kernel

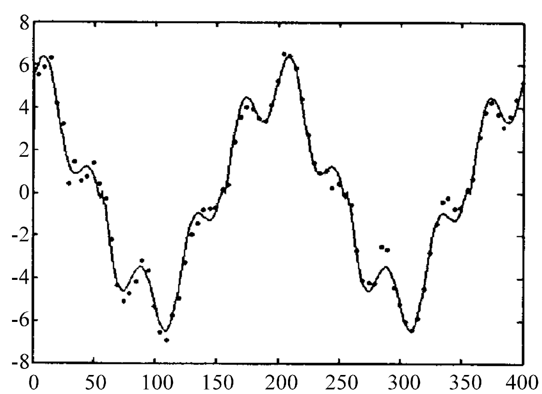

Fig. 5: The twovariable nonlinear system identification of u_1 based on Gaussian kernel

The result of the identification can be described as Table 2, Fig. 3-6. In the Fig. 3-6 real line is the real value of nonlinear system, the solid point is the identification's value. Figure 3 is the result of u_1 identification which uses the Gaussian kernel function. Figure 4 is the result of u_2 identification which uses the Gaussian kernel function. Figure 5 is the result of u_1 identification which uses the

Table 2: The twovariable nonlinear system identification result

	The parameter of kernel function	Approximation error (u_1)	Approximation error (u_2)
Gaussian kernel	$\sigma = 4$	0.0317	0.0439
Morlet wavelet kernel	a = 4	0.0173	0.0126

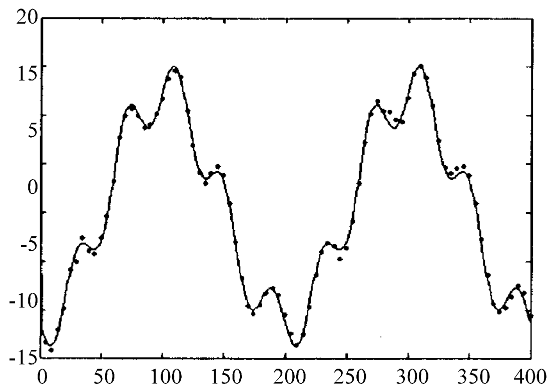

Fig. 6: The twovariable nonlinear system identification of u_2 based on Gaussian kernel

Morlet wavelet kernel function. Figure 6 is the result of u_2 identification which uses the Morlet wavelet kernel function. From these result, we can find that the Morlet wavelet kernel function not only has the capacity of non-linear mapping, but also inherit the characters of Morlet wavelet's orthonormal capacity.

CONCLUSION

For the SVM's learning method, this study proposes a new kernel function of SVM which is the Morlet wavelet kernel function. We can use this kind of kernel function to map the low dimension input space to the high dimension space. For the Morlet wavelet function, because of its horizontal floating and flexible orthonormal character, it can build the orthonormal base of $L_2(R)$ space and using this kernel function, we can approach almost any complicated functions in $L_2(R)$ space, thus this kernel function enhances the generalization ability of the SVM. At the same time, combining LS-SVM, a new method named Least squares support vector machine on Morlet wavelet kernel function is proposed and we apply it to the nonlinear system identification. Experiment shows: the Morlet wavelet kernel function is better than Gaussian kernel function.

REFERENCES

Bernhard, S. *et al.*, 1997. Comparing support vector machines with gaussian kernels to radical basis fuction classifiers. IEEE Transaction on Signal Processing, 45: 2758-2765.

Burges, C.J.C. 1998. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 2: 955-974.

Burges, C.J.C., 1999. Geometry and invariance in kernel based methods [A]. in Advance in Kernel Methods-Support Vector Learning[C]. Cambridge, MA: MIT Press, pp: 89-116.

Edgar, O. *et al.*, 1997. Training support vector machines: An application to face detection. IEEE Conference on Computer Vision and Pattern Recognition, pp: 130-136.

Mercer, J., 1909. Function of positive and negative type and their connection with the theory of integral equations [J]. Philosophical Transactions of the Royal Society of London: A, 209: 415-446.

Smola, A. *et al.*, 1998. The connection between regularization operators and support vector kernels. Neural Networks, 11: 637-649.

Suykens, J.A.K. *et al.*, 1999. Least squares support vector machine classifiers. Neural Processing Letter, 9: 293-300.

Vapnik, V., 1995. The Nature of Statistical Learning Theory. New York: Springer-Verlag, pp: 1-175.

Zhang, Q. *et al.*, 1992. Wavelet networks. IEEE Trans on Neural Networks, 3: 889-898.

Zhang, X.G., 2000. Introduction to statistical learning therory and support vector machines. Acta Automatica Snica, 26: 32-42.

Zhang, L. *et al.*, 2004. Wavelet support vector machines. IEEE Transaction on Systems, Man and Cybernetics, Part B: Cybernetics, 34: 34-39.