Multiwavelet Support Vector Machines

G. Y. Chen and W. F. Xie

Department of Mechanical & Industrial Engineering,
Concordia University,
1455 De Maisonneuve West, Montreal, Quebec, Canada H3G 1M8.
Email: {guang_c@cs.concordia.ca, wfxie@me.concordia.ca.}

Abstract

Support Vector Machines (SVM) have been developed by Vapnik and are gaining popularity because of their attractive features and promising performance in function regression and pattern classification. Recently, SVM with the scalar wavelet kernel has been developed and it has exhibited better performance in function regression and pattern recognition. In this paper, SVM with multiwavelet kernels is studied and applied to signal regression and pattern recognition. The motivation to use multiwavelets is because multiwavelets have better properties than scalar wavelets. Experiments show that this method is better than the existing SVM for function regression and pattern recognition.

Keywords: Multiwavelets, signal regression, support vector machine, pattern recognition.

1 Introduction

Since the introduction of SVM by Vapnik, it has been found to be very successful in many applications such as function regression and pattern recognition ([1] - [5]). The interesting property of SVM is that it is an approximate implementation to the structure risk minimization principal in statistical learning theory, rather than the empirical risk minimization method. An SVM uses a kernel to map the input data to a higher dimensional feature space so that the problem becomes linearly separable. The effectiveness of an SVM depends heavily on the kernel chosen. The most popular kernels include the Gaussian kernel, the polynomial kernel, the radial basis function kernel, the spline kernel, etc. Recently the scalar wavelet kernel was introduced by Zhang et al. [5] and it was found that it outperforms the Gaussian kernel for function regression and pattern recognition. Multiwavelets have many good properties that the scalar wavelets do not have, and they outperform scalar wavelets in signal denoising ([6] - [9]) and pattern recognition [10]. In this paper, we propose to use the autocorrelation of the multiwavelet functions to develop SVM kernels. Experimental results show that our multiwavelet kernels outperform other kernels for function regression and pattern recognition.

The organization of this paper is as follows. Section 2 proposes the multiwavelet SVM. Section 3 applies the multiwavelet SVM to function regression and pattern recognition. And finally section 4 gives the conclusions and future work to be done.

2 Multiwavelet SVM

Multiwavelets are generalization of scalar wavelets. Multiwavelet basis uses translations and dilations of $M \geq 2$ scaling functions $\{\phi_k(x)\}_{1 \leq k \leq M}$ and M mother wavelet functions $\{\psi_k(x)\}_{1 \leq k \leq M}$. If we write $\Phi(x) = (\phi_1(x),\ \phi_2(x),\ \ldots,\ \phi_M(x))^T$ and $\Psi(x) = (\psi_1(x),\ \psi_2(x),\ \ldots,\ \psi_M(x))^T$, then we have

$$\Phi(x) = \sqrt{2} \sum_{k=0}^{L-1} H_k \Phi(2x - k), \qquad (1)$$

and

$$\Psi(x) = \sqrt{2} \sum_{k=0}^{L-1} G_k \Phi(2x - k). \qquad (2)$$

where $\{H_k\}_{0 \leq k \leq L-1}$ and $\{G_k\}_{0 \leq k \leq L-1}$ are $M \times M$ filter matrices. As an example, for $M = 2$, $L = 4$, we give the most commonly used GHM multiwavelets developed by Geronimo, Hardin and Massopust [11]. Let

$$H_0 = \begin{pmatrix} 3/10 & 2\sqrt{2}/5 \\ -\sqrt{2}/40 & -3/20 \end{pmatrix}, \quad H_1 = \begin{pmatrix} 3/10 & 0 \\ 9\sqrt{2}/40 & 1/2 \end{pmatrix},$$

$$H_2 = \begin{pmatrix} 0 & 0 \\ 9\sqrt{2}/40 & -3/20 \end{pmatrix}, \quad H_3 = \begin{pmatrix} 0 & 0 \\ -\sqrt{2}/40 & 0 \end{pmatrix},$$

and

$$G_0 = \begin{pmatrix} -\sqrt{2}/40 & -3/20 \\ -1/20 & -3\sqrt{2}/20 \end{pmatrix}, \quad G_1 = \begin{pmatrix} 9\sqrt{2}/40 & -1/2 \\ 9/20 & 0 \end{pmatrix},$$

$$G_2 = \begin{pmatrix} 9\sqrt{2}/40 & -3/20 \\ -9/20 & 3\sqrt{2}/20 \end{pmatrix}, \quad G_3 = \begin{pmatrix} -\sqrt{2}/40 & 0 \\ 1/20 & 0 \end{pmatrix}.$$

then the two functions $\phi_1(x)$ and $\phi_2(x)$ can be generated via (1). Similarly, the two mother multiwavelet functions $\psi_1(x)$ and $\psi_2(x)$ can be constructed by (2). Multiwavelets have some advantages in comparison to scalar wavelets. For example, such features as short support, orthogonality, symmetry, and higher order of vanishing moments, are known to be important in signal processing. A scalar wavelet cannot possess all these properties at the same time, but multiwavelets can.

Smola et al. proved that a translation invariant kernel $K(\mathbf{x}, \mathbf{x}') = K(\mathbf{x} - \mathbf{x}')$ is an admissible support vector (SV) kernel if and only if its Fourier transform is non-negative [12]. This is satisfied by the scalar wavelet kernel [5]:

$$\begin{aligned} K(\mathbf{x}, \mathbf{x}') &= \prod_{i=1}^{N} h\left(\frac{x_i - x_i'}{a}\right) \\ &= \prod_{i=1}^{N} \left(cos\left(1.75 \times \frac{x_i - x_i'}{a}\right) e^{\left(-\frac{(x_i - x_i')^2}{2a^2}\right)}\right). \end{aligned}$$

where N is the dimension of the input feature vector. It is claimed in [5] that SVM with the scalar wavelet kernel outperforms the Gaussian kernel in function regression and pattern recognition. Recently multiwavelets have been found to have many advantages over scalar wavelets in signal denoising ([6] - [9]) and pattern recognition [10]. Therefore, we would like to investigate multiwavelet kernels for SVM. Since the two GHM multiwavelet functions $\psi_1(x)$ and $\psi_2(x)$ are not admissible SV kernels, the autocorrelation, $COR(\psi_k)$, of the two multiwavelet functions are used to construct two SVM multiwavelet kernels:

$$K_1(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^{N} COR\left(\psi_1\left(\frac{x_i - x_i'}{a}\right)\right) + \prod_{i=1}^{N} COR\left(\psi_2\left(\frac{x_i - x_i'}{a}\right)\right)$$

and

$$K_2(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^{N} \left(COR\left(\psi_1\left(\frac{x_i - x_i'}{a}\right)\right) \times COR\left(\psi_2\left(\frac{x_i - x_i'}{a}\right)\right)\right)$$

where N is the dimension of the input feature vector and $COR(\psi_k) = \int_{-\infty}^{+\infty} \psi_k(v)\psi_k(t+v)dv$ is the autocorrelation of the multiwavelet function $\psi_k(.), k = 1, 2$. Fig. 1 shows two GHM multiwavelet functions, their autocorrelations, and two SVM multiwavelet kernels. Our experiments show that these two multiwavelet kernels outperform the scalar wavelet kernel and the Gaussian kernel for signal regression. The following theorem states that these two multiwavelet kernels are admissible SV kernels.

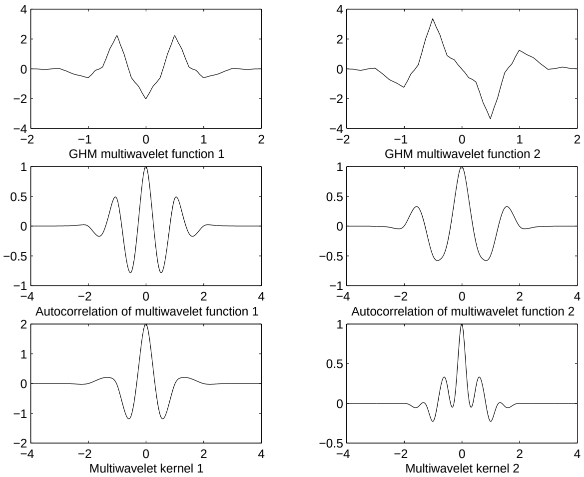

Figure 1: The two GHM multiwavelet functions, their autocorrelation and the two multiwavelet kernels.

Theorem: The multiwavelet kernels $K_1(\mathbf{x}, \mathbf{x}')$ and $K_2(\mathbf{x}, \mathbf{x}')$ defined above are admissible SV kernels.

Proof: It is easy to show that the Fourier transform of the autocorrelation function of $\psi_k(.), k = 1, 2$, is equal to the power spectrum $|F(\psi_k)|^2 \geq 0$ [13]. Therefore, $COR(\psi_k(\frac{x_i - x_i'}{a}))$ are admissible SV kernels. We also know that the sum and product of two admissible kernels are still admissible SV kernels. So it is clear that the multiwavelet kernels $K_1(\mathbf{x}, \mathbf{x}')$ and $K_2(\mathbf{x}, \mathbf{x}')$ defined above are admissible SV kernels.

3 Applications of Multiwavelet SVM

In this section, we apply the multiwavelet SVM to two important applications: function regression and pattern recognition.

3.1 Function Regression

Experiments are carried out by modifying the MATLAB SVM code developed by S. Gunn [14]. For comparison, experiments are conducted on two signal regression problems by using the two multiwavelet kernels, the scalar wavelet kernel, and the Gaussian kernel, respectively. The Gaussian kernel is defined as $K(\mathbf{x}, \mathbf{x}') = e^{(-\beta||\mathbf{x}-\mathbf{x}'||^2)}$, where $\beta > 0$ is a parameter chosen by the user. Two signals are used in our experiments:

$$f_1(x) = \begin{cases} -2.186x - 12.864, & -10 \leq x < -2 \\ 4.246x, & -2 \leq x < 0 \\ 10e^{-0.05x-0.5} sin((0.03x + 0.7)x), & 0 \leq x \leq 10. \end{cases}$$

and

$$f_2(x) = \begin{cases} 0, & -10 \leq x < -7.5 \\ 1/3, & -7.5 \leq x < -5.0 \\ 2/3, & -5.0 \leq x < -2.5 \\ 1, & -2.5 \leq x < 2.5 \\ 2/3, & 2.5 \leq x < 5.0 \\ 1/3, & 5.0 \leq x < 7.5 \\ 0, & 7.5 \leq x \leq 10. \end{cases}$$

The two signals are shown in Fig. 2. We have a uniformly sampled examples of 200 points, 25 of which are taken as training examples and others testing examples. Table 1 lists the parameters used and the regression mean square errors (MSE) for the two multiwavelet kernels, the scalar wavelet kernel, and the Gaussian kernel, respectively. We adopt the parameter $a = 1$ for the scalar wavelet kernel and $\beta = 1$ for the Gaussian kernel. The constant C and ϵ are chosen as ∞ and 0.05 for all the experiments. All these parameters are chosen the same way as [5]. We select $a = 4$ for our two multiwavelet kernels. It is found that both of the two multiwavelet kernels are better than the scalar wavelet kernel and the Gaussian kernel for signal regression. Especially, our multiwavelet kernel K_2 is better than our multiwavelet kernel K_1 in our experiments.

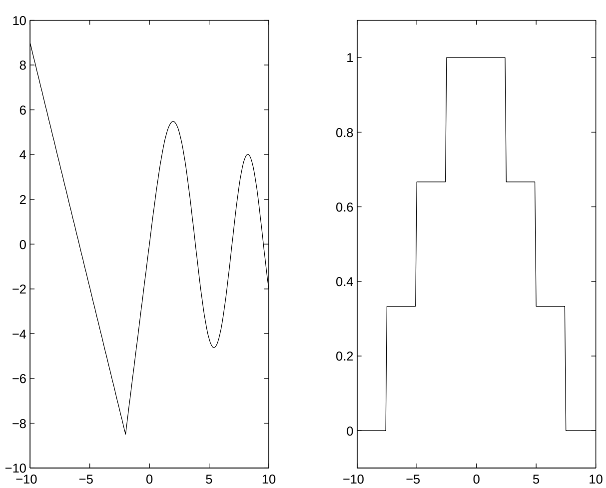

Figure 2: The 1D functions to be approximated.

We also test our multiwavelet kernels for a two dimensional function defined as

$$f(x, y) = (x^2 - y^2)sin(\frac{x}{2})$$

over the domain $[-10, 10] \times [-10, 10]$. We take 36 points as the training examples, and 1681 points as the testing examples. Fig. 3 show the original function to be approximated. Table 2 lists the approximation errors for different kernels. It is clear that the multiwavelet kernel 2 performs the best in term of approximation error.

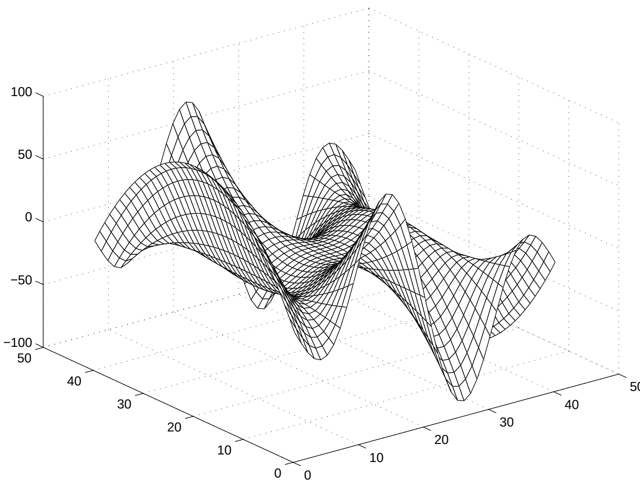

Figure 3: The 2D function to be approximated.

3.2 Pattern Recognition

In this section, we first give an invariant feature extraction descriptor, and then apply it to pattern recognition by using multiwavelet SVM.

3.2.1 Feature Extraction with Dual-Tree Complex Wavelets

It is well known that the ordinary discrete wavelet transform is not shift invariant because of the decimation operation during the transform. A small shift in the input signal can cause very different output wavelet coefficients. This is the main limitation of wavelet in pattern recognition. One way of overcomimg this is to do the wavelet transform without decimation. The drawback of this approach is that it is computationally inefficient, especially in multiple dimensions.

Kingsbury ([15], [16]) introduced a new kind of wavelet transform, called the dual-tree complex wavelet transform, that exhibits approximate shift invariant property and improved angular resolution. The success of the transform is because of the use of filters in two trees, a and b. He proposed a simple delay of one sample between the level 1 filters in each tree, and then the use of alternate odd-length and even-length linear-phase filters. As he pointed out that there are some difficulties in the odd/even filter approach. Therefore, he proposed a new

$Q-shift$ dual-tree [17] where all the filters beyond level 1 are even length. The filters in the two trees are just the time-reverse of each other, as are the analysis and reconstruction filters. The new filters are shorter than before, and the new transform still satisfies the shift invariant property and good directional selectivity in multiple dimensions. As shown later, this dual-tree complex wavelet can be sucessfully used in invariant feature extraction for pattern recognition.

We propose a novel descriptor that employs the dual-tree complex wavelet transform and multiwavelet SVM. In order to eliminate the translation variance, we move the centroid of the pattern to the center of the pattern image. Also, we normalize the pattern so that it fits into a 32×32 image. Since the dual-tree complex wavelet has the properties of shift invariance and good directional selectivity in 2D, we perform the 2D dual-tree complex wavelet on the normalized pattern and use the features at different resolution scales as a feature vector in order to train and test the SVM. There should exist two sets of data: training dataset and testing dataset. Since the dual tree complex wavelet coefficients have real and imaginary parts, we take magnitude of the complex number and use this magnitude value as our features.

The steps of the descriptor for each pattern can be summarized as follows:

1. Move the pattern centroid to the center of the pattern image.
2. Scale the pattern so that it fits exactly into a 32×32 matrix.
3. Perform the 2D dual-tree complex wavelet transform on the normalized pattern.
4. Train the multiwavelet SVM with the extracted feature vectors from the training dataset.
5. Test the multiwavelet SVM in order to get the recognition rates.

The good property of the multiwavelet, and the approximate shift invariant property of the dual-tree complex wavelet and its good directional selectivity in 2D guarantee that the new method will be better than existing methods in pattern recognition.

3.2.2 Experiments

Experiments are done on 800 training samples and 400 testing samples. Half of the training and testing samples is handwritten numeral 4 and the other half is numeral 9. Due to the large variation in the writing of these two numerals, sometimes it is quite difficult to distinguish 4 and 9. They are treated as similiar patterns in this paper. These pattern images are already segmented and isolated. Within each pattern image there exists only one pattern on an uniform background. Each original pattern is represented by 32×32 pixels. The pattern is first normalized so that it is translation- and scale-invariant, and then the dual-tree complex wavelet is applied on the normalized pattern. The wavelet coefficients of an image have multi-resolution representation of the original image. The coarse resolution wavelet coefficients normally represent the global shape of the image, while the fine resolution coefficients represent the details of the image. Our SVM code is obtained by modifying S. Gunn's Matlab code [14]. The wavelet kernel is defined as $K(\mathbf{x}, \mathbf{x}') = \prod_i \psi(\frac{x_i - x'_i}{a})$, where a is a parameter chosen by the user and $\psi(u) = cos(1.75u)e^{-u^2/2}$. We adopt the parameter $a = 4$ for the scalar wavelet kernel and $a = 12$ for the multiwavelet kernel. These parameter values are chosen by the widely used cross-validation technique ([18], [19]). The constant C is chosen as 100 for all the experiments. We use the dual-tree complex wavelet features at different resolution scales in our experiments. Table 3 lists the recognition rates for the multiwavelet kernel and the scalar wavelet kernel by using the scalar wavelet features and the dual-tree complex wavelet features. The highest recognition rate for the proposed method is 98.50% by using the dual-tree wavelet feature D_3. The highest recognition rate for the method that uses the scalar wavelet kernel and wavelet features is 94.25% by using feature D_1. It can be seen that the proposed method outperforms the method that uses the scalar wavelet kernel and wavelet features for about 4.25% in recognition rate.

4 Conclusions and Future Work

In this paper, the multiwavelet SVM has been proposed and applied to function regression and pattern recognition. The kernels used are produced by the autocorrelation of two multiwavelet functions. Experimental results show that these kernels give better results for function regression and pattern recognition. Future work will be done by applying multiwavelet SVM to other related tasks.

References

[1] V. N. Vapnik, *Statistical learning theory*, New York: Wiley, 1998.

[2] V. N. Vapnik, *The nature of statistical learning*, New York: Springer-Verlag, 1995.

[3] C. Cortes and V. N. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, pp. 273-297, 1995.

[4] Q. Song, W. J. Hu and W. F. Xie, "Robust support vector machine for bullet hole image classification," *IEEE Transactions on Systems, Man and Cybernetics - Part C*, vol. 32, no. 4, pp. 440-448, 2002.

[5] L. Zhang, W. Zhou and L. Jiao, "Wavelet support vector machine," *IEEE Transactions on Systems, Man, and Cybernetics - Part B*, vol. 34, no. 1, pp. 34-39, 2004.

[6] V. Strela, P. N. Heller, G. Strang, P. Topiwala, and C. Heil, "The application of multiwavelet filter banks to image processing," *IEEE Transactions on Image Processing*, vol.8, no.4, pp. 548-563, 1999.

[7] T. R. Downie and B. W. Silverman, "The discrete multiple wavelet transform and thresholding methods," *IEEE Transactions on Signal Processing*, vol. 46, no. 9, pp. 2558-2561, 1998.

[8] T. D. Bui and G. Y. Chen, "Translation-invariant denoising using multiwavelets," *IEEE Transactions on Signal Processing*, vol. 46, no. 12, pp. 3414-3420, 1998.

[9] G. Y. Chen and T. D. Bui, "Multiwavelet denoising using neighbouring coefficients," *IEEE Signal Processing Letters*, vol.10, no.7, pp.211-214, 2003.

[10] G. Y. Chen, T. D. Bui and A. Krzyzak, "Contour-based handwritten numeral recognition using multiwavelets and neural networks," *Pattern Recognition*, vol.36, no.7, pp.1597-1604, 2003.

[11] J. S. Geronimo, D. P. Hardin, and P. R. Massopust, "Fractal functions and wavelet expansions based on several scaling functions," *Journal of Approximation Theory*, vol. 78, pp. 373-401, 1994.

[12] A. Smola, B. Scholkopf and K.-R. Muller, "The connection between regulation operators and support vector kernels," *Neural Network*, vol. 11, pp. 637-649, 1998.

[13] K. R. Castleman, *Digital Image Processing*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632, 1979.

[14] S. Gunn, "Support vector machines for classification and regression," University of Southampton, Southampton, U.K., Image Speech and Intelligent Systems (ISIS) Group, Technical Report, May 1998.

[15] N. G. Kingsbury, "The dual-tree complex wavelet transform: a new efficient tool for image restoration and enhancement," *Proc. EUSIPCO'98*, Rhodes, Sept. 1998, pp. 319-322.

[16] N. G. Kingsbury, "Shift invariant properties of the dual-tree complex wavelet transform," *Proc. IEEE ICASSP'99*, Phoenix, AZ, March 1999.

[17] N. G. Kingsbury, "A dual-tree complex wavelet transform with improved orthogonality and symmetry properties," *Proc. IEEE ICIP*, Vancouver, Sept. 11-13, 2000.

[18] T. Joachims, "Estimating the generalization performance of a SVM efficiently," *Proc. 17th Int. Conf. Machine Learning*, San Fransisco, CA, 2000.

[19] M. Kearns and D. Ron, "Algorithmic stability and sanity-check bounds for leave-one-out cross validation," *Proc. Tenth Conf. Comput. Learning Theory*, New York, pp. 152-162, 1997.

Kernel	*Parameter*	*Approximation error* (f_1)	*Approximation error* (f_2)
Multiwavelet kernel 1	$a = 4, C = \infty, \epsilon = 0.05$	0.0918	0.0492
Multiwavelet kernel 2	$a = 4, C = \infty, \epsilon = 0.05$	**0.0820**	**0.0487**
Scalar wavelet kernel	$a = 1, C = \infty, \epsilon = 0.05$	0.1100	0.0506
Gaussian kernel	$\beta = 1, C = \infty, \epsilon = 0.05$	0.1942	0.0730

Table 1: Regression error for signal regression by using different kernels.

Kernel	*Parameter*	*Approximation error*
Multiwavelet kernel 1	$a = 6, C = \infty, \epsilon = 0.05$	3.4201
Multiwavelet kernel 2	$a = 6, C = \infty, \epsilon = 0.05$	**2.9449**
Scalar wavelet kernel	$a = 3, C = \infty, \epsilon = 0.05$	3.8028
Gaussian kernel	$\beta = 1, C = \infty, \epsilon = 0.05$	8.1546

Table 2: Regression error for 2D function regression by using different kernels.

Kernel Function	*Feature*	*Parameter Value*	*Scale*			
			D_1	D_2	D_3	D_4
Multiwavelet	*Dual-tree complex wavelet*	$C = 100, a = 12$	93.50	98.25	**98.50**	85.75
Scalar wavelet	*Scalar wavelet*	$C = 100, a = 4$	**94.25**	90.50	81.75	50.50

Table 3: The recognition rates (%) of the proposed method and the method by using the scalar wavelet kernel and scalar wavelet features.