

A Prediction Approach for Multichannel EEG Signals Modeling Using Local Wavelet SVM

Minfen Shen, *Member, IEEE*, Lanxin Lin, Jialiang Chen, and Chunqi Q. Chang, *Member, IEEE*

Abstract—Accurate modeling of the multichannel electroencephalogram (EEG) signal is an important issue in clinical practice. In this paper, we propose a new local spatiotemporal prediction method based on support vector machines (SVMs). Combining with the local prediction method, the sequential minimal optimization (SMO) training algorithm, and the wavelet kernel function, a local SMO-wavelet SVM (WSVM) prediction model is developed to enhance the efficiency, effectiveness, and universal approximation capability of the prediction model. Both the spatiotemporal modeling from the measured time series and the details of the nonlinear modeling procedures are discussed. Simulations and experimental results with real EEG signals show that the proposed method is suitable for real signal processing and is effective in modeling the local spatiotemporal dynamics. This method greatly increases the computational speed and more effectively captures the local information of the signal.

Index Terms—Electroencephalogram (EEG) signal, local prediction method, support vector machine (SVM), wavelet kernel.

I. INTRODUCTION

MANY biomedical signals contain useful information that is directly related to clinical applications. As an important category of medical signals, electroencephalogram (EEG) signals can be measured by placing many electrodes at various points on the subject's scalp. The epileptiform EEG signal includes abnormal brain electrical activity that is produced from the epileptic seizure and imposed onto the normal EEG background. To identify the region of the brain that generates the seizure activity, it is required to monitor the seizure onset and investigate the seizure pattern. EEG plays a more and more important role in the study of brain mechanism and clinical manifestations of brain diseases, with the development of signal processing technology. It is a challenge to effectively extract the useful information from multichannel EEG signals to better understand the brain functional states for various clinical applications. Many methods have been proposed to process the EEG signals for different purposes of practical applications.

More and more studies have shown that the brain is a complex nonlinear system [1]. Chaos became a promising method of EEG signal modeling and prediction [2], [3]. Chaos is based

Manuscript received June 30, 2009; revised October 20, 2009. First published February 25, 2010; current version published April 7, 2010. This work was supported in part by the National Natural Science Foundation of China under Grant 60271023 and Grant 60571066. The Associate Editor coordinating the review process for this paper was Dr. Salvatore Baglio.

M. Shen, L. Lin, and J. Chen are with the Department of Electronic Engineering, Shantou University, Shantou 515063, China (e-mail: mfshe@stu.edu.cn).

C. Chang is with the Department of Electrical and Electronic Engineering, University of Hong Kong, Pokfulam, Hong Kong.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIM.2010.2040905

on the theory of a deterministic nonlinear system that generates paths of evolution with very complex ways and appearances [4]. The seemingly random appearance of the EEG signal is, in fact, caused by the nonlinear interaction of a deterministic dynamical system. The significance of regarding EEG as chaotic is that we can use a deterministic nonlinear dynamic system to accurately model the EEG signals. Several dynamical reconstruction techniques have been developed for modeling the EEG and other biomedical signals [5], [6]. In practical applications, the model trained by observed signals can be applied to identify abnormal states of the system by detecting abrupt changes in the prediction error of the ongoing signal. In general, the prediction is one-step-ahead prediction. The importance of prediction using measured data is that it thereby provides a model for the data. On the other hand, the violation of the model can be detected, while violations often happen when the system is in abnormal states, for example, when there is epileptic seizure in the brain. Therefore, the epileptic seizures can be detected by chaotic-model-based prediction using measured EEG signals.

Chaos theory provides a way to model and predict the EEG and other signals in practice. However, it cannot make full use of all available information contained in the multichannel EEG signal as chaos is basically expressed as a temporal phenomenon. In other words, the spatial structural information of the multichannel EEG signals is ignored. As a matter of fact, the spatial structure of multichannel EEG signals is quite complex and spatially correlated in various areas of the scalp [7], [8]. To improve the performance of the multichannel EEG signal analysis, both the temporal and spatial aspects must be considered. Using the technique of reconstructing a chaotic system from the measured time series, the spatiotemporal chaotic predictor is expected to extract the information of the spatiotemporal chaotic structure.

The basis of dynamical system modeling is to reconstruct a vector space that is equivalent to the original state space of a system from a scalar time series. It is stated in the Takens embedding theorem that the reconstruction can be made by time delay embedding from a sequence of observations of the dynamical system if either the dynamics or the measurement function is generic in the sense that it couples all degrees of freedom of the system. It must be stated that: 1) this result is only an asymptotic result in the sense that it is true only when we get all the information about the dynamics of the system, i.e., when the number of samples in the observed signal goes to infinity, so that, in practice, when only finite number of samples are available, the reconstruction is only an approximation; and 2) the reconstructed space is not the original space but equivalent to the original space in the sense that the attractor

in these two spaces are equivalent, implying that we may have many different reconstructed spaces that are equivalent but not identical. Predictions about the system or the observations (in our case, the multichannel EEG signals) can be then made in the reconstructed state space by various signal processing methods, for example, a learning machine. According to the statistical learning theory, Vapnik proposed the support vector machine (SVM), which has widely been used in practical signal modeling and prediction in recent years [9], [10]. However, a global application of SVM to the whole system may have very high computational complexity since, in this case, the general direct solution of SVM needs to optimize a large quadratic problem [11], and it may not be globally optimal due to the nonstationarity of the system. Instead, a local prediction method is computationally more efficient and may provide better results for extracting the local dynamic characteristics of the EEG signals, as demonstrated in our preliminary study [12]. To evaluate the applicability of the proposed approach, the universal approximation capability (UAC) of the local reconstruction model will also be studied in this paper.

The purpose of this paper is to develop a novel spatiotemporal dynamical method for multichannel EEG signal prediction based on our preliminary study [12]. We propose a new spatiotemporal predictor, called the local wavelet SVM based on the sequential minimal optimization training algorithm (local SMO-WSVM), which combines the local prediction approach, the sequential minimal optimization (SMO) training algorithm, and the wavelet kernel to directly model the spatiotemporal chaotic signals, with applications in predicting the multichannel EEG signals.

The remainder of this paper is organized as follows. In Section II, the reconstruction of spatiotemporal dynamic signals from the measured time series is presented. Section III discusses the details of the nonlinear modeling procedures. In Section IV, simulations are carried out to evaluate the performance of the proposed local model. The experimental results with real EEG signals are illustrated in Section V. Finally, the concluding remarks are given in Section VI.

II. RECONSTRUCTION OF SPATIOTEMPORAL DYNAMICS

Let $\{s_i(\mathbf{a}_l)\}$ be a spatiotemporal signal, where \mathbf{a}_l ($l = 1, \dots, L$) represents spatial location, which is an k -dimensional vector, and $s_i(\mathbf{a}_l)$ is the 1-D state variable at spatial location \mathbf{a}_l and time $i = 1, \dots, n$. For example, when \mathbf{a}_l is a 1-D location, the entire state of the spatiotemporal signal process can then be represented by an $n \times L$ matrix [8]

$$S = [s_1, \dots, s_n]^T = \begin{pmatrix} s_1(1), & s_1(2), & \dots, & s_1(L) \\ s_2(1), & s_2(2), & \dots, & s_2(L) \\ \dots & \dots & \dots & \dots \\ s_n(1), & s_n(2), & \dots, & s_n(L) \end{pmatrix} \quad (1)$$

where $s_i(l)$ denotes the state value of the spatiotemporal system at time i and spatial site l . Let the observed spatiotemporal signal be

$$x_i(\mathbf{a}_l) = s_i(\mathbf{a}_l) + n_i(\mathbf{a}_l) \quad (2)$$

where $n_i(\mathbf{a}_l)$ is the measurement noise.

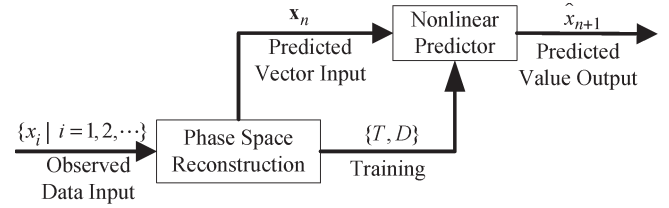


Fig. 1. Structure of spatiotemporal chaotic signal modeling.

If the spatiotemporal signal is an observation of states of a deterministic nonlinear system, for example, the multichannel EEG signal, it is needed to reconstruct (approximately) the equivalent phase space of the system from the observed spatiotemporal signal by time delay embedding according to the Takens embedding theorem. Using phase space reconstruction, we can find out the evolution rule of the chaotic attractor of the nonlinear deterministic system so that the observed signal can be modeled and described. As a result, the Takens embedding theorem provides a new vision and idea for studying the time series.

Using the multichannel EEG signal as an example, suppose the state evolution of the underlying brain dynamical system is described by a d -dimensional mapping of the form

$$\mathbf{s}_{i+1} = \varphi(\mathbf{s}_i) \quad (3)$$

where $\mathbf{s}_i \in \mathcal{R}^d$ is the state at time i , and $\varphi: \mathcal{R}^d \rightarrow \mathcal{R}^d$ is a diffeomorphism that has a dissipative chaotic attractor of dimension D_c . The collected EEG time series can be expressed by the measurement equation $x_i = h(\mathbf{s}_i)$, where $h: \mathcal{R}^d \rightarrow \mathcal{R}$ is a smooth function. The time-delay coordinate is defined by

$$\begin{aligned} \Psi_{(\varphi, h)}(\mathbf{s}) &= (h(\mathbf{s}), h(\varphi(\mathbf{s})), \dots, h(\varphi^{m-1}(\mathbf{s}))) \\ &= (x_0, x_1, \dots, x_{m-1}). \end{aligned} \quad (4)$$

According to the Takens embedding theorem [13], if $m \geq 2D_c + 1$, then we can reconstruct the chaotic dynamical system from the observed time series. More exactly, the evolution of the points $\Psi(\mathbf{s}_i) \rightarrow \Psi(\mathbf{s}_{i+1})$ in the reconstruction space follows that of the underlying dynamics φ . In other words, the collected EEG signals can be modeled by a nonlinear prediction function $\hat{x}_i = f(x_{i-1})$, where x_{i-1} denotes the vector $(x_{i-\tau}, x_{i-2\tau}, \dots, x_{i-m\tau})$, τ is the time delay, and \hat{x}_i is the predicted value of x at time i .

Fig. 1 shows the procedure of modeling the spatiotemporal chaotic signal, which is described as follow.

- 1) Phase space reconstruction: Reconstruct the system from the collected time series $\{x_i, i = 1, 2, \dots\}$ as $\{\mathbf{x}_i, i = (m-1)\tau, (m-1)\tau + 1, \dots\}$.
- 2) Training: Let the set of reconstructed vectors $T = \{\mathbf{x}_i | i = (m-1)\tau, (m-1)\tau + 1, \dots, n-1\}$ and its output $D = \{x_i | i = (m-1)\tau + 1, (m-1)\tau + 2, \dots, n\}$ be the training set of the nonlinear prediction model.
- 3) One-step-ahead prediction: Input the target vector $\mathbf{x}_n = (x_n, x_{n-\tau}, \dots, x_{n-(m-1)\tau})$ to the trained nonlinear prediction model and get the predicted value \hat{x}_{n+1} .

III. NONLINEAR MODELING

In this section, the main objective is to establish a highly efficient spatiotemporal chaotic predictor. Neural networks (NNs) and SVMs are two commonly used methods in signal modeling. SVMs, which are motivated by statistical learning theory, are regarded as one of the best machine learning methods. SVMs are based on the structural risk minimization principle so that they can minimize the data fitting error and reduce the upper bound of the generalization error at the same time, thus increasing the generalization ability of the model. First, the SVM solves the overlearning problem of NN, which employs the empirical risk minimization principle. Second, training an SVM is equivalent to solving a linearly constrained convex quadratic programming (QP) problem when the employed kernel functions satisfy the Mercer condition, thereby it does not suffer the local minima problem of NN that performs the minimization of a nonlinear error function that may be nonconvex. Finally, with the help of kernel functions, the SVM can implicitly represent a nonlinear mapping (on an infinite-dimensional space) without actually doing the mapping, which avoids the dimensionality disaster and the problem of computational complexity caused by the increase of the number of hidden layers and the structure selection of NN. Consequently, SVM is chosen as the nonlinear mapping function in this paper.

A. SVM for Prediction

Assume that the finite measured data samples $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n) \in (\mathbf{x} \times R)$ were obtained from a sample set $P(\mathbf{x}, y) (\mathbf{x} \in R^m, y \in R)$, which follows a certain distribution. The regression of SVM is to find the fitting function

$$f(\mathbf{x}) = \langle \mathbf{w}, \phi(\mathbf{x}) \rangle + b \quad (5)$$

where $\langle x, y \rangle$ denotes the inner product, and $\phi(\bullet)$ is defined as the nonlinear mapping for the input space to the feature space. The main objective of SVM regression is to find the function $f(\mathbf{x})$ to fit the samples and guarantee that it has very good generalization ability. An ε -insensitive loss function $|y_i - f(\mathbf{x}_i, \mathbf{x})|_\varepsilon = \max\{0, |y_i - f(\mathbf{x}_i, \mathbf{x})| - \varepsilon\}$ is proposed by Vapnik to enhance the robustness of the regression function. Here, $\varepsilon > 0$ is a scaling constant, which means that a regression error less than ε is ignored. Then, the ε -insensitive SVM regression can be expressed as

$$\begin{aligned} \min_{\mathbf{w}, \xi_i, \xi_i^*, b} \quad & \frac{1}{2} \langle \mathbf{w}, \mathbf{w} \rangle + C \frac{1}{n} \sum_{i=1}^n (\xi_i + \xi_i^*) \\ \text{s.t.} \quad & (\langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle + b) - y_i \leq \varepsilon + \xi_i \\ & y_i - (\langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle + b) \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0 \end{aligned} \quad (6)$$

where C is a scaling constant for the tradeoff between the regression error and the model complexity. The larger C is, the smaller the number of support vectors (SVs) and the narrower the margin. There is no systematic way for choosing C , and it

is usually empirically selected. The Lagrange function of (6) is given as follows:

$$\begin{aligned} L = \quad & \frac{1}{2} \langle \mathbf{w}, \mathbf{w} \rangle + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ & - \sum_{i=1}^n \alpha_i (\langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle + b - y_i - \varepsilon - \xi_i) \\ & - \sum_{i=1}^n \alpha_i^* (y_i - \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle - b - \varepsilon - \xi_i^*) \\ & - \sum_{i=1}^n (v_i \xi_i + v_i^* \xi_i^*). \end{aligned} \quad (7)$$

Making the partial derivatives of the Lagrange function equal to zero, (6) can yield the dual optimization problem

$$\begin{aligned} \min_{\alpha, \alpha^*} \quad & \sum_{i=1}^n [\alpha_i^* (\varepsilon - y_i) + \alpha_i (y_i + \varepsilon)] \\ & + \frac{1}{2} \sum_{i,j=1}^n (\alpha_i^* - \alpha_i) (\alpha_j^* - \alpha_j) K(\mathbf{x}_i, \mathbf{x}_j) \\ \text{s.t.} \quad & \sum_{i=1}^n (\alpha_i^* - \alpha_i) = 0 \quad 0 \leq \alpha_i, \alpha_i^* \leq C/n, \\ & i = 1, \dots, n. \end{aligned} \quad (8)$$

The linear kernel, polynomial kernel, and Gaussian kernel $K(\mathbf{x}, \mathbf{y}) = \exp(-\|\mathbf{x} - \mathbf{y}\|^2/\sigma^2)$ are commonly used. The dynamic system model can be obtained by solving the dual optimization problem, and the optimum solution $\bar{\alpha} = [\bar{\alpha}_1^*, \bar{\alpha}_1, \bar{\alpha}_2^*, \bar{\alpha}_2, \dots, \bar{\alpha}_n^*, \bar{\alpha}_n]^T$ is obtained. For the input vector \mathbf{x} , the prediction can be deduced from

$$f(\mathbf{x}) = \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle + \bar{b} = \sum_{SV} (\bar{\alpha}_i^* - \bar{\alpha}_i) K(\mathbf{x}_i, \mathbf{x}) + \bar{b} \quad (9)$$

where the bias \bar{b} is given by

$$\bar{b} = \frac{1}{l_{SV}} \sum_{j=1}^{l_{SV}} \left(y_j - \sum_{i=1}^{l_{SV}} (\bar{\alpha}_i^* - \bar{\alpha}_i) K(\mathbf{x}_i, \mathbf{x}_j) \right) + \varepsilon \quad (10)$$

where l_{SV} denotes the number of SVs.

B. Local SMO-WSVM Model

The solution process of the SVM dual problem is equivalent to a QP problem with linear constraints that requires storage of a kernel function matrix with size proportional to the square of the number of training samples. Therefore, with the number of samples increased, more storage memory is needed. Even if there are only a few thousands of data points, it still requires hundreds of megabytes of memory. For short data, it may not be a problem if the algorithm is run on a modern personal computer equipped with 2 or 3 GB of main memory, but it may cause problems for long data, or the algorithm is implemented on low-end devices such as microcontrollers or field-programmable gate arrays. In addition, a large number of

matrix operations need to be carried out to solve the quadratic optimization problem associated with the SVM. In most cases, the optimization algorithm for SVM training consumes most of the computing time. When the number of samples becomes large, the time required to solve the QP problem for training SVM may become a bottleneck of practical applications of SVM. Therefore, we need to find a prediction algorithm to reduce the computational complexity of the model.

The local prediction method of chaotic series, which does not fit all the state points reconstructed from all sample data, is different from the global prediction method [14]. The local prediction method only selects a small number of state points (p points) to fit the function f at \mathbf{x}_n , where f may be either nonlinear (polynomial, rational fraction, etc.) or linear around \mathbf{x}_n . Therefore, the overall f is piecewise linear or nonlinear and can thus express the overall nonlinear characteristics of the chaotic signal.

The procedure of the local prediction method is described as follows.

- 1) According to the Takens embedding theorem, select the appropriate embedding dimension m and time delay τ and then reconstruct from a spatiotemporal chaotic series $\{x_i | i = 1, 2, \dots\}$ to an equivalent phase space as $\{\mathbf{x}_i, i = (m-1)\tau, (m-1)\tau + 1, \dots\}$. Let \mathbf{x}_n denote the target vector to be predicted and calculate the distances between the target vector \mathbf{x}_n and the preformed N state vectors as

$$d(i) = \|\mathbf{x}_i - \mathbf{x}_n\|, \quad i = (m-1)\tau, (m-1)\tau + 1, \dots \quad (11)$$

- 2) Select p reconstructed states that are close to the target vector \mathbf{x}_n , denoted as the training input vectors T_{Local} , and then find out the corresponding p training output vectors, denoted as D_{Local}

$$T_{\text{Local}} = [\mathbf{x}_{n_1}, \mathbf{x}_{n_2}, \dots, \mathbf{x}_{n_p}] \quad (12)$$

$$D_{\text{Local}} = [x_{n_1+1}, x_{n_2+1}, \dots, x_{n_p+1}] \quad (13)$$

where n_1, n_2, \dots, n_p denote the locations that are close to \mathbf{x}_n .

- 3) Use T_{Local} as the training input vectors and D_{Local} as the training output vectors to train the SVM model.
- 4) Take \mathbf{x}_n as the input to the trained SVM model, and then, the prediction value \hat{x}_{n+1} can be obtained.
- 5) Let $n \leftarrow n + 1$ and iterate steps 1–4 until all prediction values are obtained.

Using the local prediction method, the p training samples close to the target vector are used for training the SVM model at each iteration. The QP problem is obviously simplified since the size of the problem is significantly reduced in the local models, which means that the computational complexity of the SVM model is also significantly reduced. On the other hand, since only local training samples are used to train the SVM models, for nonstationary signals that have local dynamics, modeling with the local prediction method obtains better performance than the global prediction method.

Although the local prediction method has greatly reduced the computational complexity of the SVM model, the direct

solution of the QP problem in SVM training is still the main reason for limiting the efficiency of the model when the number of close points p greatly increases. Therefore, it is desired to find a fast training algorithm. Many efficient training algorithms have recently been proposed, such as the SVM light decomposition algorithm [15], the SMO algorithm proposed by Platt [16], the neighbor algorithm proposed by Kerthi [17], and the least square SVM (LS-SVM) algorithm proposed by Suykens and Vandewalle [18], where the first three algorithms are based on inequality constraints and the last one is based on equality constraints. The SMO training algorithms iteratively select two data points and optimize the objective function with respect to them, resulting in an analytic solution where the problem of solving the QP is avoided. Although the number of iterations in the SMO training algorithm is larger, each iteration is very fast so that the total computation time is greatly reduced. Another advantage of the SMO training algorithm is that there is no need to store the kernel matrix. On the other hand, the LS-SVM algorithm translates the standard SVM optimization problem with inequality constraints into an optimization problem with equality constraints, making the original QP problem into a matrix calculation, thereby greatly enhancing the speed of model training. However, comparing with the algorithm based on inequality constraints, the LS algorithm based on equality constraints cannot greatly reduce the complexity of the model. A side effect of the LS-SVM algorithm is that it also reduces the accuracy of the SVM model. Therefore, the SMO training algorithm is used in this paper to improve the speed of the local SVM model by considering both the complexity and accuracy of the model.

Combining the local SVM prediction method and the SMO training algorithm, an efficient local dynamic model is established. We need to further study the UAC of this model. The capability of the nonlinear approximation is achieved by the kernel mapping, which maps the data in input space into a high-dimensional feature space. Accordingly, the problem of making the model have the UAC is equal to the problem of selecting the kernel function.

Dot product, polynomial, and Gaussian kernels are the normal kernels that are generally used in many applications, and among which, the Gaussian kernel is the most popular one. Current research works show that SVMs with standard kernels, including Gaussian, polynomial, and several dot product kernels, can approximate any measurable or continuous function up to any desired accuracy [19]. In other words, these SVMs have the UAC in the continuous space.

Theoretically, wavelet decomposition emerges as a powerful tool for approximation [20]. A set of basis that can approximate arbitrary functions can be obtained through three ways: 1) translate the scaling function with a particular scale parameter; 2) translate the wavelet functions with some scale parameters; and 3) vary the scale parameter of wavelet functions and translate the scaling function with a larger scale parameter. In consequence, these SVMs with scaling/wavelet kernels probably have the UAC so that it has the theoretical advantages to investigate the SVMs with scaling/wavelet kernels (simplified as wavelet kernels). Based on the considerations aforementioned, the wavelet SVM (WSVM) model has been a hot research topic.

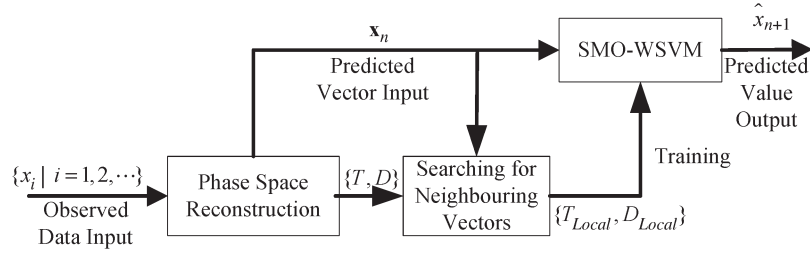


Fig. 2. Structure of local spatiotemporal chaotic signal prediction model based on the SMO-WSVM method.

The kernel that satisfies the Mercer condition is called an admissible SV kernel, which ensures the global optimality of the solutions. Similarly, translation-invariant kernels, i.e., $K(\mathbf{x}, \mathbf{x}') = K(\mathbf{x} - \mathbf{x}')$, are admissible SV kernels if they satisfy the Mercer condition. However, it is difficult to decompose the translation-invariant kernels into the product of two functions and then prove them as SV kernels. A necessary and sufficient condition for translation-invariant kernels is described as follows.

Theorem 1 [20]: A translation-invariant kernel $K(\mathbf{x}, \mathbf{x}') = K(\mathbf{x} - \mathbf{x}')$ is an admissible SV kernel if and only if the Fourier transform

$$F[K(\mathbf{w})] = 2\pi^{-N/2} \int_{R^N} \exp(-j\langle \mathbf{w}, \mathbf{x} \rangle) K(\mathbf{x}) d\mathbf{x} \geq 0 \quad (14)$$

is founded.

Theorem 2 [21]: Let $\varphi(x)$ be the wavelet function; then, the wavelet kernel function of the SVM is

$$K(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^m \varphi\left(\frac{x_i - c_i}{\sigma}\right) \varphi\left(\frac{x'_i - c'_i}{\sigma}\right) \quad \mathbf{x}, \mathbf{x}' \in R^m. \quad (15)$$

For the translation-invariant kernel, (15) can be rewritten as

$$K(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^m \varphi\left(\frac{x_i - x'_i}{\sigma}\right). \quad (16)$$

Theorem 3 [22]: The Mexican Hat wavelet kernel function that satisfies the translation-invariant characteristic

$$K(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^m \left(1 - \frac{\|x_i - x'_i\|^2}{a^2}\right) \exp\left(-\frac{\|x_i - x'_i\|^2}{2a^2}\right) \quad (17)$$

is an admissible SV kernel.

Considering the construction of the SV kernel functions of SVMs, we have theoretically proven that the wavelet kernels are the feasible SV kernels and superior to other kernels. Hereto, a nonlinear model with the UAC has been built.

For the reasons provided earlier, according to the ideas of a local prediction method [13], [14], a local SMO-WSVM model has been proposed in this paper. More precisely, first, the local prediction method has the characteristics of small-sample training, a simplified model, and high accuracy. Second, the SMO algorithm has not only significant advantage in speed of training but also high accuracy. Finally, wavelet kernels have the theoretical advantages that have the UAC. Combining all of these advantages, an effective chaotic predictor that has the

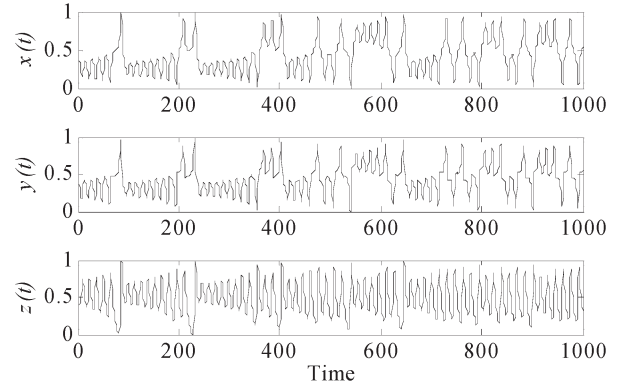


Fig. 3. Lorenz chaotic time series.

local dynamic characteristics is constructed. Fig. 2 shows the structure of the local spatiotemporal chaotic signal prediction model based on the SMO-WSVM method.

IV. SIMULATIONS

To verify the prediction performance of the local SMO-WSVM model to spatiotemporal chaotic signals, a series of data was generated from the Lorenz equation

$$\begin{cases} dx/dt = a(y - x) \\ dy/dt = cx - xz - y \\ dz/dt = xy - bz. \end{cases} \quad (18)$$

If the parameters are appropriately set, the Lorenz system can be chaotic. Here, we set the following parameters: $a = 10$, $b = 8/3$, and $c = 28$. The time t is sampled from 0 to 60 with a time step of 0.05, and the initial value is set to (1,1,1). After removing the first 201 transition points, we can get a chaotic time series with 1000 points $\{x(t)\}$, $\{y(t)\}$, $\{z(t)\}$. Then, these three chaotic series are normalized with the following equation (using $\{x(t)\}$ as an example) to make their values lie between 0 and 1:

$$x(t) = \frac{x(t) - \min(x(t))}{\max(x(t)) - \min(x(t))}. \quad (19)$$

The three chaotic time series are shown in Fig. 3.

The first 800 points are chosen as the training samples, and the following 200 points are chosen as the prediction samples. Let the embedding dimension $m = 3$ and time delay $\tau = 1$; then, the phase space is reconstructed according to the Takens embedding theorem. The backpropagation (BP) NN and the SVM are used as the nonlinear predictor, and their performance is compared. The number of hidden centers of the BP is set

TABLE I
PREDICTION ERRORS OF $\{y(t)\}$ WITH SEVEN LEARNING MACHINES

Method	MSE	Time (s)
BP	1.6075×10^{-4}	-
Global QuadProg-GSVM	8.7483×10^{-4}	2447
Local QuadProg-GSVM	2.4277×10^{-5}	14.12
Global SMO-GSVM	9.8658×10^{-6}	20.34
Local SMO-GSVM	1.4320×10^{-6}	10.06
Global SMO-WSVM	7.6602×10^{-6}	62.05
Local SMO-WSVM	1.3614×10^{-6}	10.12

to ten, and the training precision is 0.005. We compare several different SVM approaches with a combination of global and local prediction methods, the direct solution of the large-scale convex QP problem and the SMO training algorithm (“SMO” for short), Gaussian kernel (“GSVM” for short), and Mexican hat wavelet kernel (“WSVM” for short). In the training algorithms, the parameters are set as $C = 1000$ and $\varepsilon = 0.001$, with the scale parameter being $\sqrt{3}$. In the local SVM training, the number of closest data points is set as $p = 20$. We use the chaotic time series $\{y(t)\}$, which has a strong local dynamic characteristic, to test the methods.

Table I shows the prediction mean square errors (MSEs) of $\{y(t)\}$ for seven learning machines. The program was run under the Matlab platform on a DELL computer that has a Pentium(R)-4 CPU of 2.4 GHz and 512-MB random access memory. From Table I, we can see that the SVM methods, in general (except the global QuadProg-GSVM), have smaller prediction errors than the BP method. However, when we adjust the parameters to make the training error smaller, the prediction of NN would become worse. The reason is that due to the empirical risk minimization principle the NN has the overlearning problem. The performance of NN is also dependent on the size of the training set. When the training set reduces, the performance of the model will rapidly drop.

From the simulation results of various SVM approaches, the following can be observed: 1) Under the same conditions, the local methods have a faster prediction speed than the global methods. Taking the method of the direct solution of the large-scale convex QP problem, for instance, the local QuadProg-GSVM is 180 times faster than the global QuadProg-GSVM. When using the SMO training method, the prediction rate of the local methods also improves by one to five times. 2) Under the same conditions, the local methods have better performance than global methods. From Table I, we can see that, whether using the QP training method or SMO, the prediction performance of the local method is better than that of global methods for both GSVM and WSVM. 3) The performance of WSVM is better than that of the GSVM. For global methods with SMO training, the computation complexity of the global SMO-WSVM is more than twice that of the global SMO-GSVM. The reason is that the Mexican hat wavelet kernel is more complex than the Gaussian kernel. For local methods, the computation complexity is almost the same for both models.

From the aforementioned analysis, the computation complexity of the model is significantly reduced because the local

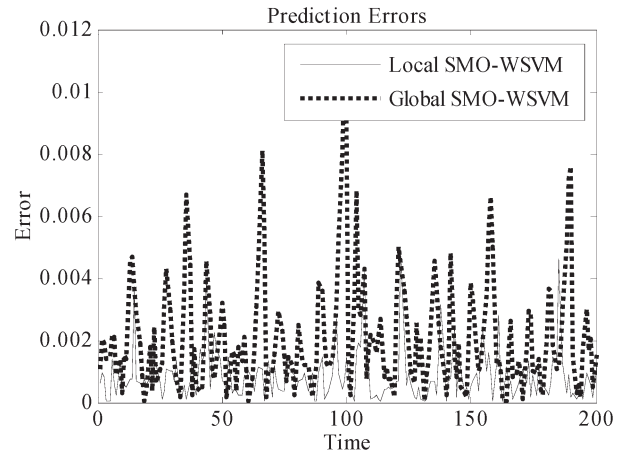


Fig. 4. One-step prediction error of chaotic time series.

method only use p samples close to the target vector for training. Moreover, the local method builds up a local dynamic model, which effectively captures the local information of the system, so it has better prediction performance than the global method to those signals that have local dynamic characteristics. Furthermore, the SMO optimal training algorithm improves the efficiency of training, making it better than the QP training algorithm in both prediction speed and prediction performance. Finally, the experiment shows that the prediction performance of SVMs with wavelet kernels is much better than the performance with Gaussian kernels.

The proposed local prediction method significantly reduces the computation complexity of the SVM model and improves the prediction performance of the dynamic signal with local characteristics. We combine the SMO training method to further improve the computation speed of the model. The prediction errors of both the local SMO-WSVM model and the global SMO-WSVM model are shown in Fig. 4. It can be seen from Fig. 4 that the local SMO-WSVM has the advantages of modeling signals with local dynamic characteristics.

V. APPLICATIONS TO REAL EEG SIGNAL

A. Real EEG Prediction

The brain is regarded as a spatiotemporal chaotic system, and different mental states correspond to different potential dynamical systems. In this section, the EEG data with both normal EEG and epileptic seizure signals are from the Set A and Set E of EEG recordings, respectively, provided by Andrzejak *et al.* [23].

The proposed method is used to model and predict the EEG signals. First, the EEG signal is normalized according to (19). Then, the first 800 points were used for training the SVM, and the following 200 points were used for testing. The embedding dimension was selected as $m = 5$, and the delay time was $\tau = 1$ in the reconstruction of the phase space according to the Takens embedding theorem. Note that selecting the delay τ and selecting the embedding dimension m are key and nontrivial steps in applying the Takens embedding theorem to recovering nonlinear dynamics. In this study, the embedding dimension is selected according to the false nearest neighbor and Cao

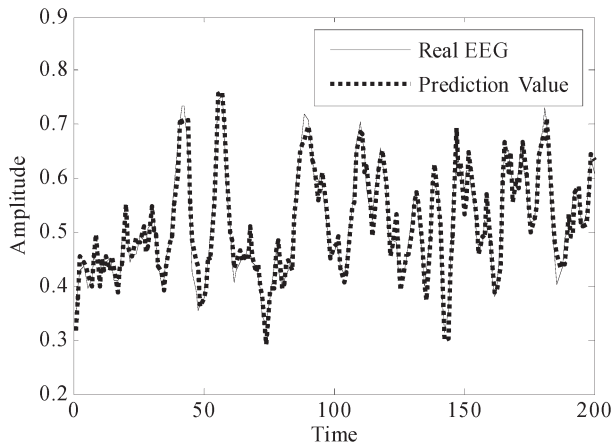


Fig. 5. Prediction result of a real EEG signal.

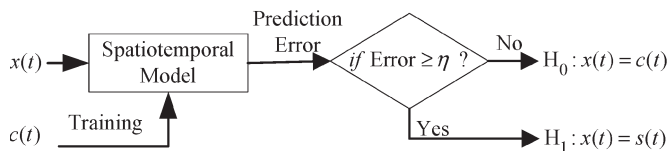


Fig. 6. Epilepsy detection framework based on the proposed model.

method, and the time delay is selected according to mutual information analysis. For more details, the readers are referred to reference [7] and the references therein. The local SMO-WSVM approach was used to predict the EEG signal, where the training parameters were set as $C = 1000$ and $\varepsilon = 0.001$. The number of close points is set as $p = 30$. Fig. 5 shows the prediction result. The MSE of the prediction is 9.0365×10^{-4} . It can be seen that the proposed model successfully and effectively recovered the EEG signal and modeled the spatiotemporal chaotic signal.

B. Epilepsy Detection Based on Local SMO-WSVM

Due to the difference in the dynamical characteristics between normal EEG signals and epilepsy seizure signals, a new method of detecting epileptic seizure based on the local SMO-WSVM model is proposed. The detection framework is shown in Fig. 6, which is described as follows.

- 1) Both normal EEG signals and epileptiform EEG signals are recorded and represented as $c(t)$ and $s(t)$, respectively.
- 2) Train the local SMO-WSVM model with normal EEG signals $c(t)$ and then take the trained model as a predictor.
- 3) The testing EEG signal is represented by $x(t)$. Take the signal $x(t)$ as the local SMO-WSVM model's input.
- 4) Predict the testing signals with the local SMO-WSVM model. Compare the prediction error with a threshold η . If the error is less than the specified threshold, set the input signal as H_0 , namely, $x(t) = c(t)$. Otherwise, the input signal is set as H_1 , namely, $x(t) = s(t)$.

To evaluate the detection performance based on the local SMO-SVM model, we choose 1200-point EEG data for this purpose. Here, the first 1000 points are normal EEG signals, and the last 200 points are the epilepsy seizure signals. The first 800 points are used to train the model, and the last 400 points

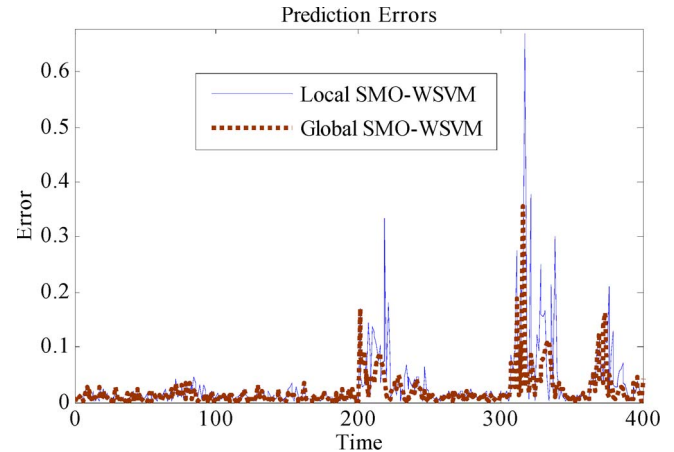


Fig. 7. (color) Prediction errors in normal and epileptic seizure EEG, the first 200 points are the prediction error of normal EEG, and the last 200 points are the prediction error of epileptic seizure EEG.

are used for detection experiment. The detecting error results based on the local SMO-WSVM model and the global SMO-WSVM model are shown in Fig. 7. It can be seen from the result that the prediction errors are obviously larger at the last 200 points than the first 200 points predicted by both models. There are mainly two reasons for this: 1) these two models, which are both trained based on normal EEG signals, can exactly describe the characteristics and evolvement of chaotic dynamical of EEG, so the detection error will be small in the nonepilepsy region; and 2) the chaotic dynamic characteristics have an obvious difference between epilepsy and normal EEG signals. Therefore, the prediction error will be larger at the epilepsy region, and the proposed model is effective for epilepsy detection. Furthermore, as shown in Fig. 7, the prediction errors with the local SMO-WSVM model in the epilepsy region is larger than that of global SMO-WSVM model, which means that the proposed model holds a good performance in epileptic seizure detection.

VI. CONCLUSION

In this paper, we have proposed a novel spatiotemporal prediction model. A spatiotemporal chaotic predictor called the local SMO-WSVM has been developed to model the real multichannel EEG time series. The proposed model can significantly reduce the complexity of the model and capture the local information of the signals more effectively in the engineering applications. Furthermore, theoretical analysis shows that the proposed method has the UAC. Both simulations and experiments using real EEG signals suggest that the presented method is much more effective in predicting spatiotemporal chaotic time series, and the local SMO-WSVM model is suitable to model the spatiotemporal dynamic system and more efficient in EEG signal processing.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments that greatly improve the presentation of this paper.

REFERENCES

- [1] N. V. Thakor and S. Ton, "Advances in quantitative electroencephalogram analysis methods," *Annu. Rev. Biomed. Eng.*, vol. 6, no. 1, pp. 453–495, 2004.
- [2] B. Joydeep, "Complexity analysis of spontaneous EEG," *Acta Neurobiol. Exp.*, vol. 60, no. 4, pp. 495–501, 2000.
- [3] X. Zhang, R. J. Roy, and E. W. Jenson, "EEG complexity as a measure of depth of anesthesia for patients," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 12, pp. 1424–1433, Dec. 2001.
- [4] H. Zhang, T. N. Zhang, J. H. Shen, and X. J. Liu, "Research on characteristics of chaos in enterprise strategic system and chaos management," in *Proc. ICMSE*, 2008, pp. 393–399.
- [5] D. M. Tumeay, P. E. Morton, D. F. Ingle, C. W. Downey, and J. H. Schnurer, "Neural network classification of EEG using chaotic preprocessing and phase space reconstruction," in *Proc. IEEE 17th Annu. Northeast Bioeng. Conf.*, 1991, pp. 51–52.
- [6] Y. W. Tao, X. Xu, and X. F. Cheng, "A blind deconvolution method of single-channel the EEG convolution mixed signals based on the linear predication error analysis," in *Proc. 2nd Int. Conf. Bioinform. Biomed. Eng.*, 2008, pp. 2083–2086.
- [7] M. F. Shen, L. X. Lin, and G. L. Chang, *Novel Coupled Map Lattice Model for Prediction of EEG Signal*. Berlin, Germany: Springer-Verlag, pt. 1, pp. 347–356.
- [8] Y. S. Xia and H. Leung, "Nonlinear spatial-temporal prediction based on optimal fusion," *IEEE Trans. Neural Netw.*, vol. 17, no. 4, pp. 975–988, Jul. 2006.
- [9] J. Smola and B. Scholkopf, *A Tutorial on Support Vector Regression*. London, U.K.: Univ. London, 1998.
- [10] J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Knowl. Discov. Data Mining*, vol. 2, no. 2, pp. 121–167, Jun. 1998.
- [11] V. P. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
- [12] M. F. Shen, J. L. Chen, and C. H. Lin, "Modeling of nonlinear medical signal based on local support vector machine," in *Proc. IEEE IMTC*, Singapore, May 5–7, 2009, pp. 675–679.
- [13] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence*. Berlin, Germany: Springer-Verlag, 1981, pp. 366–381.
- [14] J. S. Zhang, J. L. Dang, and H. C. Li, "Local support vector machine prediction of spatiotemporal chaotic time series," *Acta Physica Sinica*, vol. 56, no. 1, pp. 67–77, 2007, (in Chinese).
- [15] T. Joachims, "Making large scale SVM learning practical," in *Advances in Kernel Methods: Support Vector Machines*. Cambridge, MA: MIT Press, Dec. 1998.
- [16] J. C. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machine," in *Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA: MIT Press, 1999.
- [17] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to Platt's SMO algorithm for SVM classifier design," Dept. Mech. Prod. Eng. Nat. Univ. Singapore, Singapore, TR CD 299214, 1999.
- [18] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, Jun. 1999.
- [19] B. Hammer and K. Gersmann, "A note on the universal approximation capability of support vector machines," *Neural Process. Lett.*, vol. 17, no. 1, pp. 43–53, Feb. 2003.
- [20] A. Smola, B. Schölkopf, and K. R. Müller, "The connection between regularization operators and support vector kernels," *Neural Netw.*, vol. 11, no. 4, pp. 637–649, Jun. 1998.
- [21] L. Zhang, W. D. Zhou, and L. C. Jiao, "Wavelet support vector machine," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 1, pp. 34–39, Feb. 2004.
- [22] W. Z. Cui, C. C. Zhu, W. X. Bao, and J. H. Liu, "Least squares wavelet support vector machines and its application to nonlinear system identification," *J. Xi'an Jiaotong Univ.*, vol. 38, no. 6, pp. 563–565, 2004, (in Chinese).
- [23] R. G. Andrzejak, K. Lehnertz, C. Rieke, F. Mormann, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 64, no. 6, p. 061 907, Dec. 2001.



Minfen Shen (M'09) received the M.Eng. degree in signal processing and the Ph.D. degree in communication and information systems from the University of Science and Technology of China (USTC), Hefei, China, respectively.

After graduation, he joined the Department of Electrical and Electronic Engineering, Shantou University, Shantou, China, where he is currently a Professor. He is also an Adjunct Professor with the Department of Scientific Technology, USTC. His research interests include random signal processing, computational intelligence, biomedical signal analysis, nonlinear signal and image processing, and chaos.



Lanxin Lin received the M.S. degree in signal and information processing from the Shantou University, Shantou, China, in 2009. She is currently working toward the Ph.D. degree with the Department of Electronic Engineering, City University of Hong Kong, Kowloon, Hong Kong.

Her current research interests include nonlinear signal processing, statistical signal processing, and initial conditions estimation in coupled map lattices.



Jialiang Chen received the M.S. degree in signal and information processing from the Shantou University, Shantou, China, in 2009.

He is currently with the Department of Electronic Engineering, Shantou University. His research interests include image processing, nonlinear signal processing and shape modeling.



Chunqi Q. Chang (M'06) received the B.Sc. (Eng.) and the M.Sc. (Eng.) degrees in electronic engineering from the University of Science and Technology of China, Hefei, China, in 1992 and 1995, respectively, and the Ph.D. degree in biomedical engineering from the University of Hong Kong, Pokfulam, Hong Kong, in 2001.

Since 2002, he has been with the Department of Electrical and Electronic Engineering, University of Hong Kong, where he is currently a Research Assistant Professor. His research interests are in the areas of signal processing and machine learning with applications in biomedical engineering and bioinformatics, particularly statistical signal processing, blind signal processing, computational neuroscience and neuroengineering, brain computer interface, event-related brain potentials, and computational systems biology.

Dr. Chang is a Reviewer for *Bioinformatics*, *Briefings in Bioinformatics*, *Human Brain Mapping*, *Psychometrika*, *PROCEEDINGS OF IEEE*, *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, *IEEE TRANSACTIONS ON NEURAL NETWORKS*, *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS*, *Signal Processing*, *Digital Signal Processing*, *IET Signal Processing*, *EURASIP Journal on Applied Signal Processing*, and *Journal of Medical Systems*. He is on the Editorial Advisory Board of *The Open Spectroscopy Journal*.