Available online at www.sciencedirect.com
ScienceDirect
Journal of Sound and Vibration 311 (2008) 886–897
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

Nonlinear structural response prediction based on support vector machines

Dong Yinfeng[a,*], Li Yingmin[a], Lai Ming[b], Xiao Mingkui[a]

[a]*College of Civil Engineering, Chongqing University, Chongqing 400045, China*
[b]*Department of Science and Technology, Ministry of Construction, Beijing 100835, China*

Received 14 December 2006; received in revised form 16 July 2007; accepted 25 September 2007
Available online 8 November 2007

Abstract

A support vector machines (SVM)-based two-stage method is proposed to simulate and predict the nonlinear dynamic response of structures. In the first stage, an autoregressive moving average with exogenous input (ARMAX) model is used to represent the acceleration response as the output of a single-input single-output (SISO) system and the least square method is used to estimate the model parameters with which the linear acceleration response of the system can be simulated and predicted. Then the linear velocity and displacement are estimated using numerical integration of the predicted acceleration. In the second stage, by using the predicted linear responses (acceleration, velocity and displacement) and the excitation to construct the input vector, the SVM is used to approximate nonlinear mapping from the input vector to system output and the trained SVM can be used to simulate and predict the nonlinear dynamic response conveniently. The nonlinear dynamic responses of a Duffing oscillator and a frame structure are simulated and predicted using the proposed method as well as the neural network-based method. The results demonstrate that the SVM-based method provides superior performance in generalization and accuracy and can be a powerful tool for nonlinear system simulation and prediction.

1. Introduction

As a powerful tool for structural seismic design, control, identification, damage detection and health monitoring, nonlinear structural response analysis has become a subject of intensive study for decades. Among the various methods, time history analysis is the most frequently used numerical method. With numerous studies having been carried out on it, this method has been used successfully in the fields of civil engineering. But it is still a challenging topic because of the complexity of structural models and uncertainty in excitation. For this reason, many researchers have attempted to develop other alternative methods to circumvent the above difficulty, which gives rise to the development of the system identification-based methods. In these methods, the nonlinear mapping of the given (or known) excitation to the corresponding outputs is used to simulate and predict the nonlinear dynamic response. As for nonlinear system identification,

*Corresponding author. Tel.: +86 23 6512 1991; fax: +86 23 6512 3511.
E-mail address: yhvhson@cta.cq.cn (D. Yinfeng).

0022-460X/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsv.2007.09.054

numerous methods, excellent and comprehensive reviews of works on nonlinear system identification and its applications in structural response simulation, health monitoring and loss estimation were given in Refs. [1–4].

In recent years, considerable attention has been focused on the neural network which has been proven to be an efficient nonlinear approximator [5–8]. In extensive studies, the potential power of neural network was represented; meanwhile, some difficulties were also emphasized [9–20]. First, the generalization ability of neural network which was very sensitive to the number of training samples, network topology and learning algorithms was still a critical issue. Second, in some studies, to simulate and predict a concerned response, other measured responses were also required besides the excitation, i.e. to simulate and predict the acceleration response the measured velocity and displacement was required as well as the excitation, which made the neural network suitable for identification other than prediction of nonlinear systems. Last, for nonlinear systems, the simulation and prediction of acceleration response was not studied as extensively as the displacement and it was really a difficult problem. Furthermore, recent developments in vibration-based structural damage detection and health monitoring have indicated the increasing importance of acceleration response and an efficient method for simulating and predicting the acceleration response is thought to be useful for future research.

Compared to the neural network, the support vector machine (SVM) developed by Vapnik [21] recently is another powerful tool for general classification and regression problems which provides the advantage in generalization and global optimization. As a classifier it was used widely for structural damage detection [2,22,23], but as a nonlinear approximator its application was only reported in a few studies [24–26].

To address the above problems and attracted by the excellent performance of the SVM, we propose a SVM-based two-stage method to simulate and predict the acceleration response of nonlinear systems and hope it to be an efficient tool for structural damage detection and health monitoring.

The theory of neural network has been reported in extensive literatures, thus only the background of SVM regression is summarized in the following section. Then the details of the proposed method are illustrated and the results for the simulation and prediction of a Duffing oscillator and a 5-story frame structure are demonstrated and discussed. As a counterpart to the proposed method, a three-layered neural network is also used to simulate and predict the acceleration response of the nonlinear systems in the second stage. Lastly, some notes on SVM for nonlinear system simulation and prediction are remarked.

2. Theory of SVM regression

Originally, SVM was designated to solve classification problems. However, with the introduction of Vapnik's ε-insensitive loss function, SVM has been extended to solve nonlinear regression estimation problems successfully [21].

In SVM regression the basic idea is to map the input data x into a high-dimensional feature space via a nonlinear mapping Φ and to do linear regression in this space. The regression model is defined as $y = f(x) + e$, where x and y are, respectively the input and output, $f(x)$ is the linear regression function defined in the high-dimensional feature space, e is the independent random error. Given n sampling (input–output) pairs $G = \{(x_i, y_i), i = 1,2, \ldots, n\}$, SVM approximates the linear regression function $f(x)$ given by

$$f(x) = \omega\Phi(x) + b, \tag{1}$$

where $\Phi(x)$ is the high-dimensional feature space which is nonlinearly mapped from the input space x. Eq. (1) is also called the decision function. The coefficients ω and b are estimated by minimizing

$$R_{\mathrm{reg}}(C) = C\frac{1}{n}\sum_{i=1}^{n} L_{\varepsilon}(y_i, f(x_i)) + \frac{1}{2}\|\omega\|^2, \tag{2}$$

$$L_{\varepsilon}(y, f(x)) = \begin{cases} y - f(x)| - \varepsilon & \text{for } |y - f(x)| \geqslant \varepsilon, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Eq. (2) is called the regularized risk function where the first term $C(1/n)\sum_{i=1}^{n} L_{\varepsilon}(y_i, f(x_i))$ is the empirical error (risk), and the second term $\frac{1}{2}\|\omega\|^2$, on the other hand, is the regularization term. Eq. (3) is the ε-insensitive loss

function which provides the advantage of enabling one to use sparse data points (sampling pairs) to represent the decision function given by Eq. (1). C is the regularized constant and it determines the trade-off between the empirical risk and the regularization term. Increasing the value of C will result in the relative importance of the empirical risk with respect to the regularization term to grow. ε is called the tube size and it is equivalent to the approximation accuracy placed on the training sampling pairs.

To obtain the estimations of ω and b, Eq. (2) is transformed to the primal function by introducing the positive slack variables ξ_i and ξ_i^* as follows:

$$\begin{aligned} &\text{Minimize} && R(\omega, \xi^{(*)}) = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ &\text{subjected to} && y_i - \omega\Phi(x_i)b \leqslant \varepsilon + \xi_i, \\ &&& \omega\Phi(x_i) + b - y_i \leqslant \varepsilon + \xi_i^*, \\ &&& \xi^{(*)} \geqslant 0. \end{aligned} \tag{4}$$

Introducing Lagrange multipliers and exploiting the optimal constraints, Eq. (1) can be represented in explicit form as

$$f(x, a_i, a_i^*) = \sum_{i=1}^{n}(a_i - a_i^*)K(x, x_i) + b. \tag{5}$$

In Eq. (5), a_i and a_i* are the Lagrange multipliers. They satisfy $a_i \times a_i^* = 0$, $a_i \geqslant 0$, and $a_i^* \geqslant 0$, and are obtained by maximizing the dual function of Eq. (4) in the following form:

$$R(a_i, a_i^*) = \sum_{i=1}^{n} y_i(a_i - a_i^*) - \varepsilon\sum_{i=1}^{n}(a_i + a_i^*) - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(a_i - a_i^*)(a_j - a_j^*)K(x_i, x_j) \tag{6}$$

with the constraints $\sum_{i=1}^{n}(a_i - a_i^*) = 0$, $0 \leqslant a_i \leqslant C$ and $0 \leqslant a_i^* \leqslant C$.

Based on the Karush–Kuhn–Tucker (KKT) conditions of quadratic programming, only a certain number of coefficients $(a_i - a_i^*)$ in Eq. (5) will assume non-zero values. The data points associated with them have approximation errors equal to or larger than ε and are referred to as support vectors. These are the data points lying on or outside the ε-bound of the decision function. According to Eq. (5), it is evident that support vectors are the only elements of the data points that are used in determining the decision function as the coefficients $(a_i - a_i^*)$ of other data points are all equal to zero. Generally, the larger the ε, the fewer the number of support vectors and thus the sparser the representation of the solution. However, a larger ε can also depreciate the approximation accuracy placed on the training points. In this sense, ε is a trade-off between the sparseness of the representation and closeness to the data.

$K(x, x_i)$ is defined as the kernel function. The value of the kernel is equal to the inner product of two vectors x_i and x_j in the feature space $\Phi(x_i)$ and $\Phi(x_j)$, that is, $K(x, x_i) = \Phi(x_i) * \Phi(x_j)$. The elegance of using the kernel function is that one can deal with feature spaces of arbitrary dimensionality without having to compute the map $\Phi(x)$ explicitly. Any function satisfying Mercer's condition can be used as the kernel function [21]. The typical examples of kernel function are the polynomial kernel $K(x,y) = (x * y + 1)^d$ and the Gaussian kernel $K(x, y) = \exp(-1/\delta^2(x-y)^2)$ where d is the degree of polynomial kernel and δ^2 is the bandwidth of the Gaussian kernel. The kernel parameter should be carefully chosen as it implicitly defines the structure of the high-dimensional feature space $\Phi(x)$ and thus controls the complexity of the final solution. From the implementation point of view, training SVM is equivalent to solving a linearly constrained quadratic programming with the number of variables twice as that of the training data points. The sequential minimal optimization algorithm proposed by Smola [27] is thought to be very effective in training SVM. Once the SVM is trained, it can be used for simulation and prediction purpose according to Eq. (5).

Different from most of the traditional neural networks which implement the empirical risk minimization principle, SVM implements the structural risk minimization principle to minimize an upper bound of the generalization error rather than minimize the training error. This induction principle is based on the fact that the generalization error is bounded by the sum of the training error and a confidence interval term that depends on the Vapnik–Chervonenkis (VC) dimension. Based on this principle, SVM can achieve an optimum

network structure by striking a right balance between the empirical error and the VC-confidence interval, which eventually results in better generalization performance than the traditional neural networks. Another merit of SVM is that the training process is equivalent to solving a linearly constrained quadratic programming, which means the solution of SVM is unique, optimal and absent from local minima, unlike the training process of the traditional neural networks which requires nonlinear optimization thus running the danger of getting stuck in a local minima.

3. The two-stage method

Considering the excellent performance of SVM, we propose a SVM-based two-stage method to simulate and predict the acceleration response of nonlinear systems. In the first stage, an autoregressive moving average with exogenous input (ARMAX) model is used to represent the acceleration response as the output of a single-input single-output (SISO) system, and the least square method [28] is used to estimate the model parameters with which the linear acceleration response of the nonlinear system can be simulated and predicted based on the initial conditions. Then the linear velocity and displacement components can be estimated using numerical integration of the predicted acceleration. In the second stage, using the predicted linear responses (acceleration, velocity and displacement) and the excitation to construct input vector, SVM is used to approximate nonlinear mapping from the input to output and the trained SVM can be used to simulate and predict the nonlinear dynamic response conveniently. Fig. 1 presents the diagram of the method and it can be summarized in the following steps:

1. Use ARMAX (p, q) model to represent the concerned acceleration response of the nonlinear system given by

$$y(k) = \sum_{i=1}^{p} \phi_i y(k-i) + \sum_{i=1}^{q} \theta_i x(k-i) + x(k) + e_a(k), \tag{7}$$

   in which the variables $y(k)$ and $x(k)$ are the output and input of the system sampled at the time instant $t = k\,\Delta t$ with Δt as the sampling interval. ϕ_i and θ_i are the autoregressive and moving average coefficients, respectively. $e_a(k)$ is measurement noise.
2. Use least square method [28] to estimate the model parameters ϕ_i and θ_i.

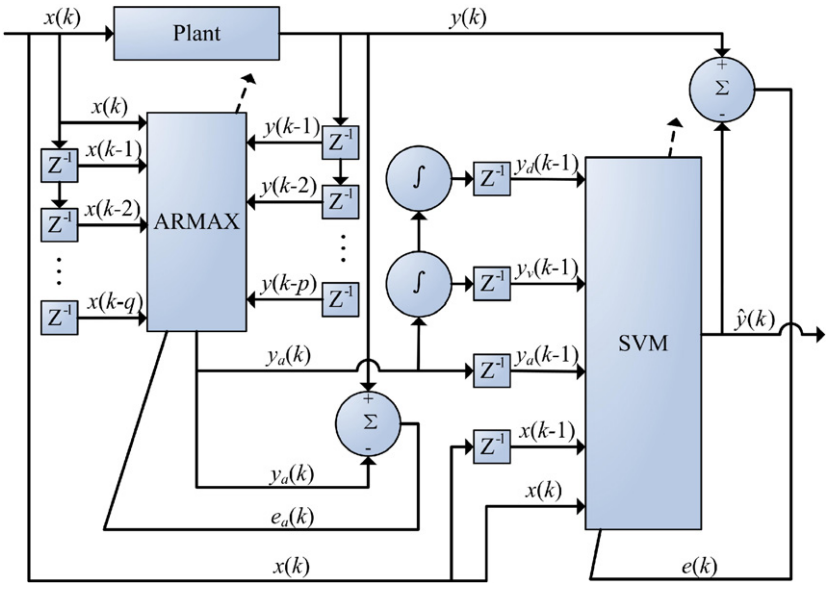

Fig. 1. Diagram of the SVM-based two-stage method.

3. The linear acceleration response $y_a(k)$ is predicted using ϕ_i, θ_i and the initial conditions. Then, the linear velocity $y_v(k)$ and displacement $y_d(k)$ can be estimated using numerical integration of $y_a(k)$.
4. Use $y_a(k-1)$, $y_v(k-1)$, $y_d(k-1)$, $x(k-1)$ and $x(k)$ to construct input vector $\mathbf{X}(k)$ corresponding to the output $y(k)$ and train SVM with the input–output pairs $G\{(\mathbf{X}(k), y(k)), i = 1,2, \ldots, n\}$. Then, the acceleration response of the nonlinear system can be simulated and predicted with the input vector $\mathbf{X}(k)$ and the trained SVM according to Eq. (5).

It has been shown by Pandit [29] and Anderson et al. [30,31] that an ARMAX model of order $(2m, 2m-1)$ is equivalent to a system with m degrees of freedom (dof), thus, the order of ARMAX model can be determined according to the dof of the system.

4. Results and discussions

To demonstrate the performance of the proposed method, two examples are given here, i.e. the acceleration responses of a Duffing oscillator and a 5-story frame structure are simulated and predicted. In both examples, the acceleration responses to the acceleration record El Centro (1940, N–S) in Fig. 2(a) are first used to train the SVM and then with the trained SVM the acceleration responses to an artificial wave in Fig. 2(b) are predicted to explore the prediction and generalization performance of the method. As a counterpart to the proposed method, a three-layered neural network (NN) is also used in the second stage. The numbers of neurons for the three layers are 30, 30 and 1, respectively. The activation functions are linear functions for the first and the third layers and radial basis function (RBF) for the second layer. For the proposed method, only the 500 sampling pairs for the first 10 s are used to train the SVM, while for the NN-based method all the 2000 sampling pairs are used to train the NN.

4.1. Example 1: Duffing oscillator

As a classic nonlinear model, the Duffing oscillator has been studied extensively in the literature. The motion equation is given by $m\ddot{y} + c\dot{y} + k_1 y + k_3 y^3 = -m\ddot{x}(t)$ where $\ddot{y}$, $\dot{y}$ and y are the acceleration, velocity and displacement responses, respectively and the input acceleration is denoted as $\ddot{x}(t)$. The system parameters used in this example are chosen as follows: mass $m = 1\,\mathrm{kg}$ damping coefficient $c = 0.316\,\mathrm{N\,s\,m^{-1}}$, linear stiffness coefficient $k_1 = 10\,\mathrm{N\,m^{-1}}$ and cubic stiffness coefficient $k_3 = -160\,\mathrm{N\,m^{-3}}$. With zero initial conditions, its responses are calculated by applying the fourth-order Runge–Kutta method using MATLAB

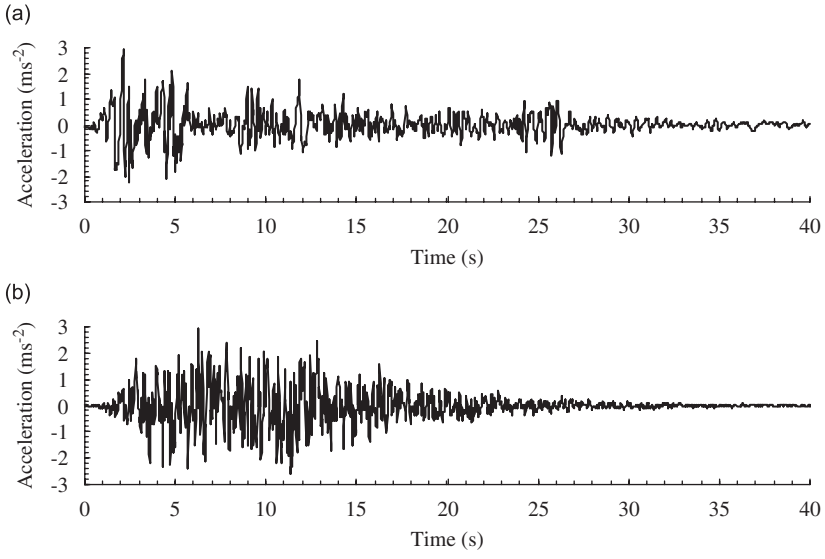

Fig. 2. Accelerograms of the two inputs: (a) El Centro (1940, N–S) and (b) artificial wave.

software and the time step is set as 0.02 s. In the first stage of the SVM- and the NN-based methods, the ARMAX (2, 1) model is used to predict the linear response components for the Duffing oscillator is a sdof system.

It is evident that for a Duffing oscillator the restoring force $F(y)$ is a third-order polynomial function of the displacement y, i.e. $F(y) = k_1 y + k_3 y^3$. Fig. 3 illustrates the displacement, restoring force and restoring force vs. displacement curves of the Duffing oscillator corresponding to the inputs in Fig. 2 with the same system parameters as mentioned above. We can see that there is a slight difference between the two curves due to the distinctive properties (amplitude and frequency properties) of the two inputs, and the maximum displacement response to El Centro (1940, N–S) is lager than that to the artificial wave. This difference is what we expect, for this difference can be exploited to evaluate the prediction and generalization ability of the two methods.

The actual and predicted accelerations corresponding to various inputs and methods are shown in Fig. 4. From visual inspection, it can be seen that both methods provide pretty good performance in generalization and prediction as a whole. The SVM-based method slightly outperforms the NN-based method in prediction. The generalization ability for the SVM-based method is obviously better, for it uses only 500 sampling pairs to train the SVM, while in the NN-based method 2000 sample pairs are used to train the NN.

Fig. 5 presents the statistical distribution of the actual and the predicted acceleration responses in Fig. 4. We can see that the distribution patterns of predicted accelerations are similar to those of actual accelerations, but the results of the SVM-based method show a far better agreement than the results of the NN-based method. In

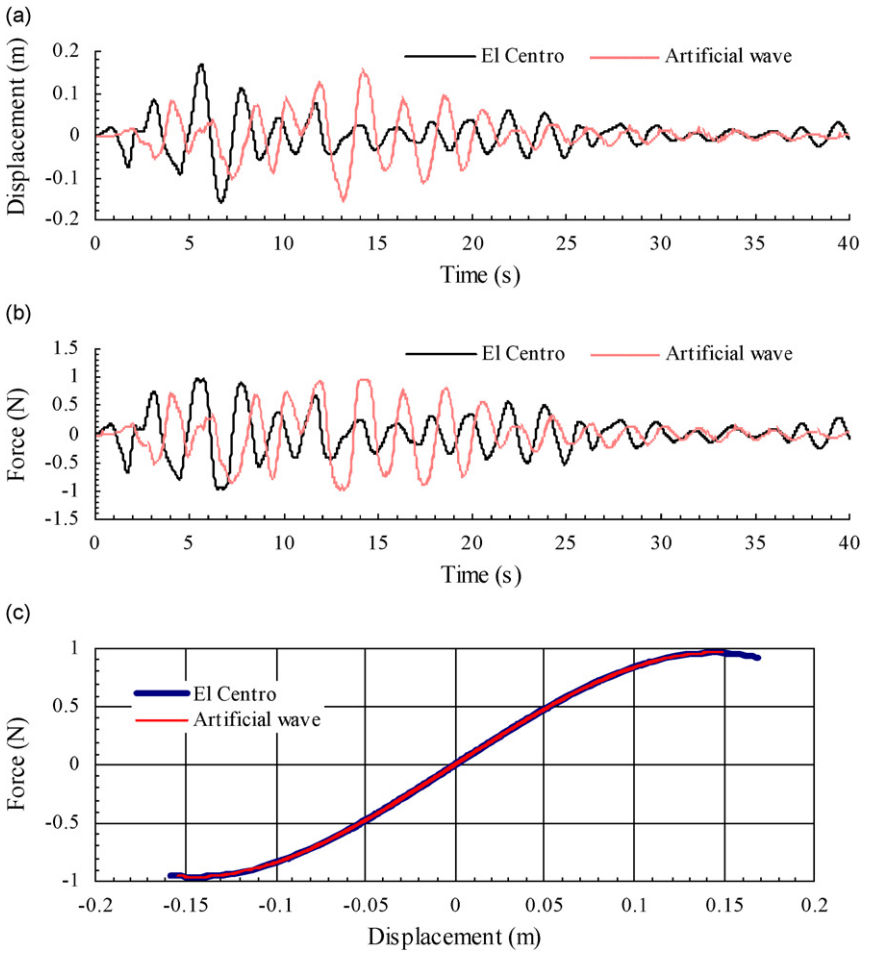

Fig. 3. Responses of the Duffing oscillator corresponding to the two inputs given in Fig. 2 with the same system parameters: (a) displacement; (b) restoring force; (b) restoring force vs. displacement curves.

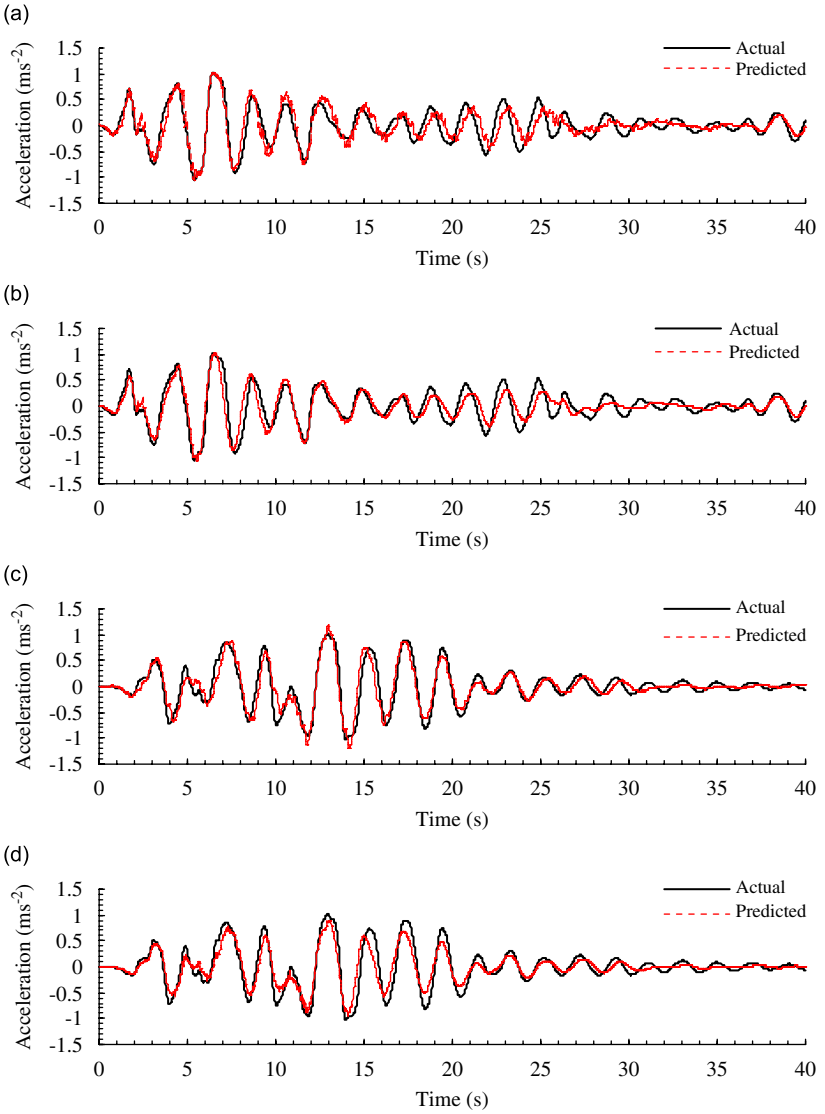

Fig. 4. Actual and predicted acceleration responses of the Duffing oscillator corresponding to various inputs and methods: (a) El Centro (1940, N–S) and SVM-based method; (b) El Centro (1940, N–S) and NN-based method; (c) artificial wave and SVM-based method; (d) artificial wave and NN-based method.

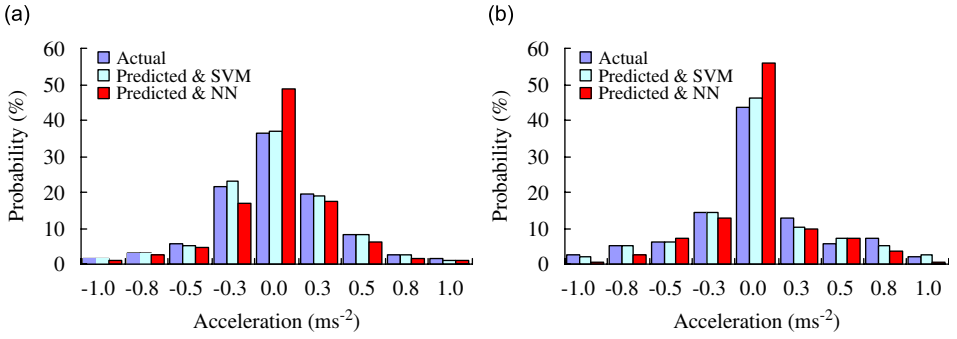

Fig. 5. Statistical distribution of the actual and predicted acceleration responses in Fig. 4: (a) responses corresponding to El Centro (1940, N–S) and (b) responses corresponding to artificial wave.

NN cases, the maximum probability is relatively higher than in cases of SVM and actual response which indicates that the acceleration response predicted using NN-based method is somewhat smaller than the actual response, and the prediction ability of NN-based method is poorer than SVM-based method. In cases of El Centro (1940, N–S), the maximum probability is smaller in comparison to the artificial wave, which is mainly because the acceleration response to artificial wave is much smaller than to El Centro (1940, N–S) after 20 s as can be seen from Fig. 4.

The cumulative energy defined as the cumulative sum of the squared acceleration is an important quantity to describe the non-stationary properties of structure response, thus it is used in this study to evaluate the agreement between the actual and the predicted accelerations. The cumulative energy curves corresponding to the actual and the predicted accelerations in Fig. 4 are presented in Fig. 6 which shows more clearly than Figs. 4 and 5 that the SVM-based method performs better than the NN-based method in prediction and generalization. For the NN-based method, the agreement in Figs. 6(b, d) is somewhat worse than in Figs. 4(b, d) and 5(a, b), while for the SVM-based method the agreement is quite well in all cases.

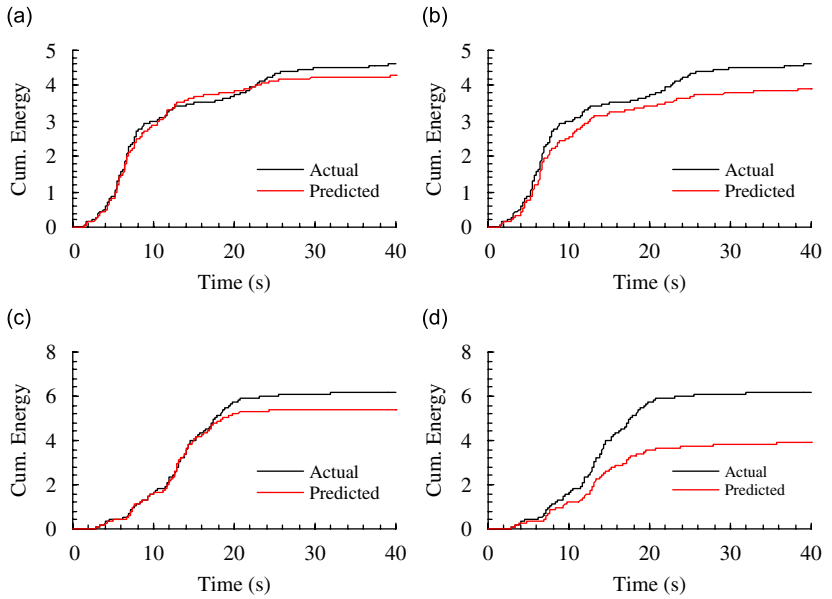

Fig. 6. Cumulative energy of the actual and predicted acceleration responses in Fig. 4: (a) El Centro (1940, N–S) and SVM-based method; (b) El Centro (1940, N–S) and NN-based method; (c) artificial wave and SVM-based method; (d) artificial wave and NN-based method.

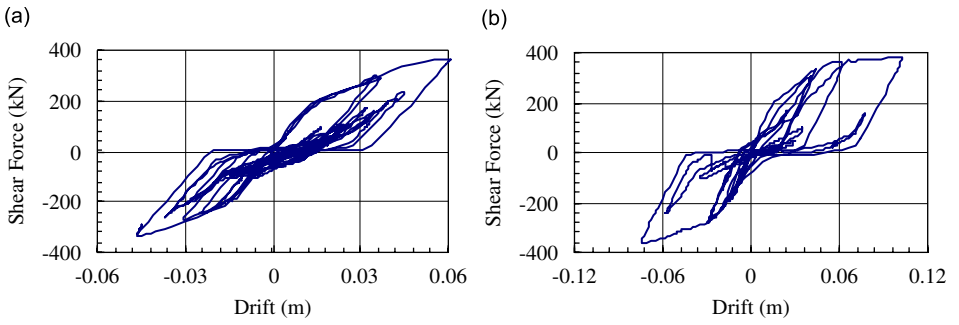

Fig. 7. Story shear force vs. story drift curves of the first floor corresponding to the inputs in Fig. 2: (a) El Centro (1940, N–S) and (b) artificial wave.

4.2. Example 2: 5-story frame structure

In this example, we use the proposed method to simulate and predict the acceleration response of a reinforced concrete frame [32]. The frame is 5-story, 3-bay plane structure. The height for the first floor is 5.5 and 3.3 m for the second–fifth floors, the width for each bay is 6 m. The inelastic responses of the frame are analyzed using the computer program IDARC 2D [33] with the trilinear hysteretic model. It is found that the first floor is the weak story of the structure, thus the acceleration responses of the first floor are selected as the target to perform simulation and prediction. In the first stage of the proposed method, the ARMAX model of order (10, 9) is used for the system is a 5-dof system.

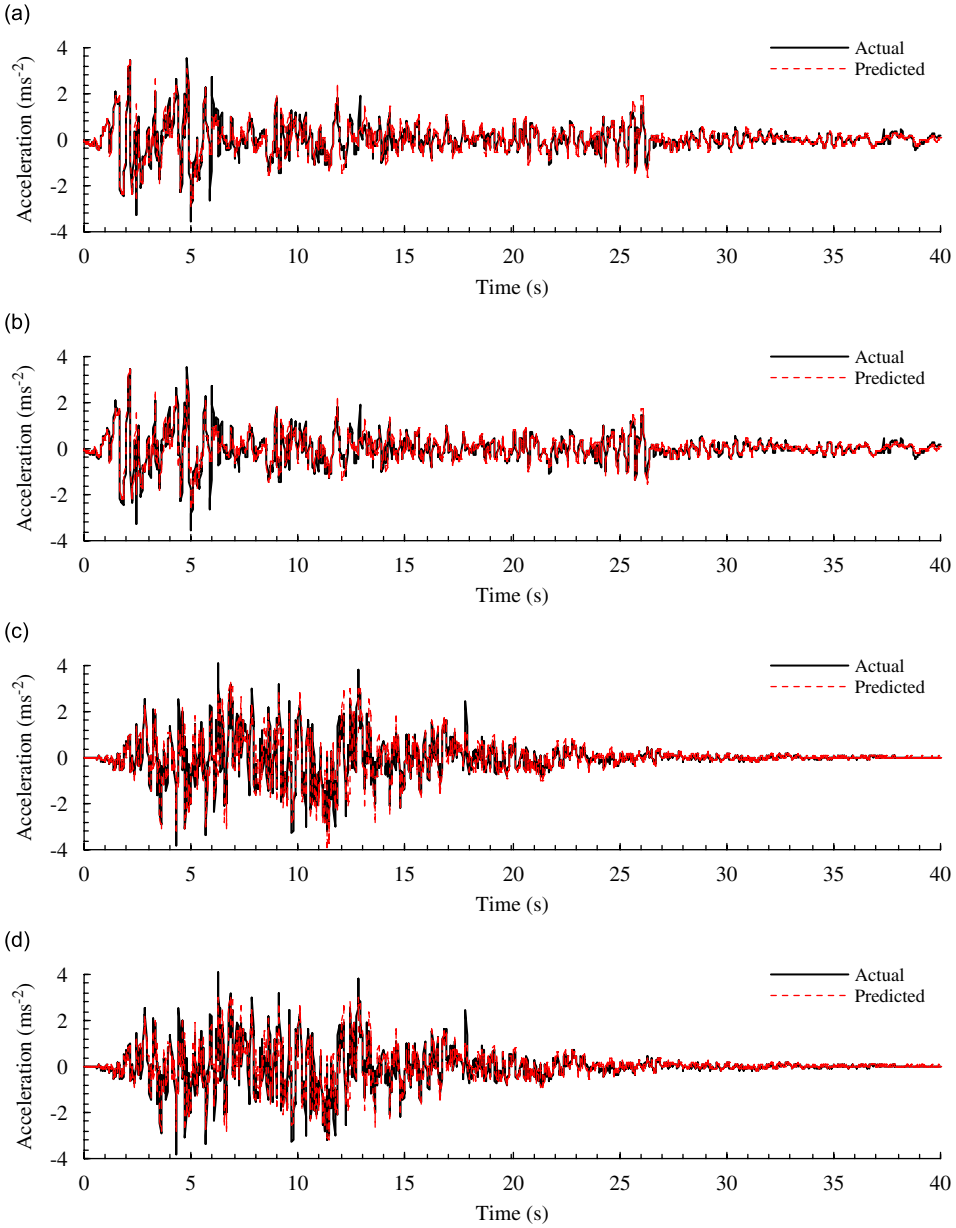

Fig. 8. Actual and predicted acceleration responses of the first floor corresponding to various inputs and methods: (a) El Centro (1940, N–S) and SVM-based method; (b) El Centro (1940, N–S) and NN-based method; (c) artificial wave and SVM-based method; (d) artificial wave and NN-based method.

This example is undoubtedly more complicated than the first one due to the hysteresis properties of the structure. Fig. 7 shows the story shear force vs. story drift curves of the first floor corresponding to the inputs in Fig. 2. The two curves are obviously different as we expect, and the maximum displacement is observed in case of artificial wave. But with the same inputs, in case of the first example, maximum displacement is observed in case of El Centro (1940, N–S). This is mainly because the frequency properties of the two inputs and the dynamic properties (e.g., natural frequency) of the two systems are different. More details regarding this point can be found in Ref. [32].

Fig. 8 demonstrates the actual and the predicted accelerations of the first floor corresponding to various inputs and methods. From visual inspection, it is clear that the prediction results of both methods are in good agreement with the actual responses and there is no significant difference between them.

Fig. 9 shows the statistical distribution of the actual and the predicted acceleration responses of the first floor. It also presents excellent agreement between the prediction results and the actual responses, but also

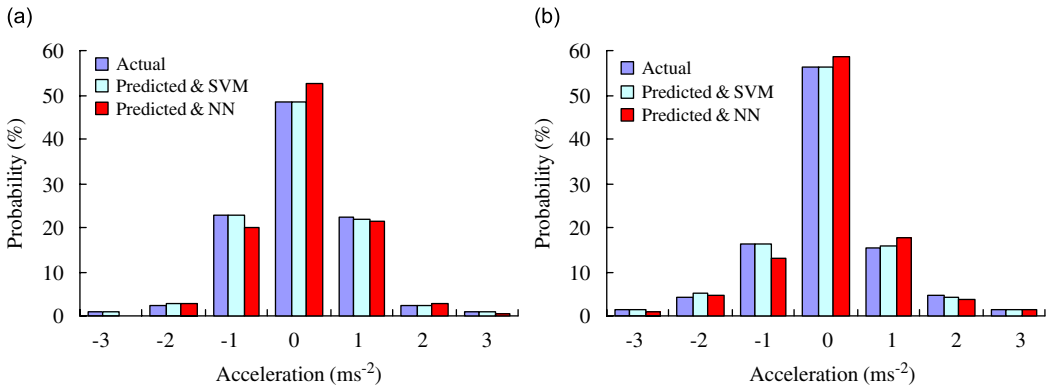

Fig. 9. Statistical distribution of the actual and predicted acceleration responses in Fig. 8: (a) responses corresponding to El Centro (1940, N–S) and (b) responses corresponding to artificial wave.

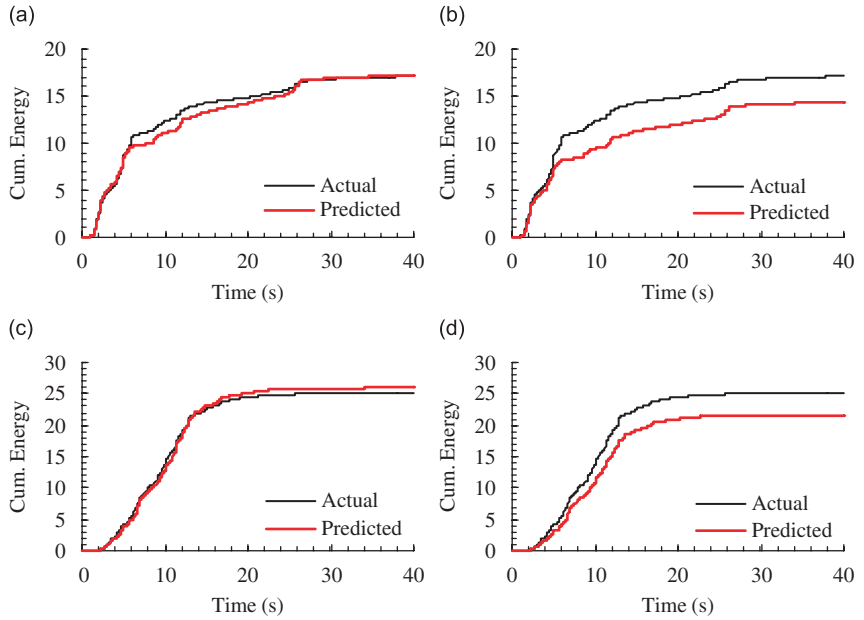

Fig. 10. Cumulative energy of the actual and predicted acceleration responses in Fig. 8: (a) El Centro (1940, N–S) and SVM-based method; (b) El Centro (1940, N–S) and NN-based method; (c) artificial wave and SVM-based method; (d) artificial wave and NN-based method.

indicates that the SVM-based method performs relatively better than the NN-based method. In NN cases, as same as what is shown in Fig. 5, the maximum probability is relatively higher than in cases of SVM due to the poorer prediction ability of NN-based method. And in cases of El Centro (1940, N–S), the maximum probability is smaller in comparison to the artificial wave, which is mainly because the acceleration response to artificial wave is much smaller than to El Centro (1940, N–S) after 20 s due to the difference of the two inputs.

The cumulative energy curves corresponding to the actual and the predicted accelerations are illustrated in Fig. 10. It shows, more clearly than Figs. 8 and 9, that the SVM-based method performs much better than the NN-based method. The results of the SVM-based method are very close to the actual results, but the results of the NN-based method are about 5–15 percent smaller than the actual results.

5. Conclusions

In this study, a SVM-based two-stage method is proposed to simulate and predict the acceleration response of nonlinear structures. The method is applied to a Duffing oscillator and a 5-story frame structure and the results show that the method is efficient and its generalization and prediction abilities are quite well. This is mainly because that SVM converts the nonlinear regression in low-dimensional space to the linear regression in high-dimensional feature space and carries out the regression estimation under the structural risk minimization principle. Although we do not compare the proposed method with other methods suggested in the literature, the prediction precision and generalization ability of the method is sufficient for practical purpose. We find that the cumulative energy, as an important quantity to describe the non-stationary properties of structure response, is a sensitive index to evaluate the performance of prediction methods. It is also to be noted that for nonlinear systems they may present the possibility of multiple solutions, jump phenomena and chaotic properties due to the bifurcations of equilibrium positions of them, and the initial conditions and the nature of excitation will determine which of these solutions will represent the actual response of them. Thus in the first stage of the proposed method the linear response components should be predicted based on the actual initial conditions, otherwise prediction errors will appear in the first several cycles of response. For example, in cases of non-zero initial conditions the zero initial conditions assumption will yield considerable errors in the first 2 or 3 cycles of response. Though the proposed method has the potential ability to deal with multiple solutions problems, the training samples should be selected as properly as to cover the nature of possible excitation thoroughly. Really, it is an important issue being studied. Lastly, for SVM is performed in high-dimensional space, the efficient training algorithm of SVM for a large scale of data is an important issue for future research.

Acknowledgments

This work was supported by National Natural Science Foundation of China under Grant nos. 50008017 and 50578167.

References

[1] G. Kerschen, K. Worden, A.F. Vakakis, J.C. Golinval, Past, present and future of nonlinear system identification in structural dynamics, *Mechanical Systems and Signal Processing* 20 (2006) 505–592.

[2] H. Sohn, C.R. Farrar, F.M. Hemez, D.D. Shunk, D.W. Stinemates, B.R. Nadler, A review of structural health monitoring literature: 1996–2001, *Los Alamos National Laboratory LA-13976-MS*, 2003.

[3] K.A. Porter, J.L. Beck, J.Y. Ching, J. Mitrani-Reiser, M. Miyamura, A. Kusaka, T. Kudo, K. Ikkatai, Y. Hyodo, Real-time loss estimation for instrumented buildings, *Earthquake Engineering Research Laboratory EERL 2004-08*, 2004.

[4] J.Y. Ching, J.L. Beck, K.A. Porter, R. Shaikhutdinov, Real-time Bayesian state estimation of uncertain dynamical systems, *Earthquake Engineering Research Laboratory EERL 2004-01*, 2004.

[5] T.P. Chen, H. Chen, Approximations of continuous functions by neural networks with application to dynamic system, *IEEE Transactions on Neural Networks* 4 (1993) 910–918.

[6] T.P. Chen, H. Chen, Approximation capability to functions of several variables, nonlinear functionals, and operators by radial basis function neural networks, *IEEE Transactions on Neural Networks* 6 (1995) 904–910.

[7] B. Irie, S. Miyake, Capabilities of three-layered perceptrons, *Proceedings of IEEE International Conference on Neural Networks*, Vol. 1, 1988, pp. 641–648.
[8] K. Hornik, Approximation capabilities of multilayer feedforward networks, *Neural Networks* 4 (1991) 251–257.
[9] S. Chen, S.A. Billings, Neural networks for nonlinear dynamic system modeling and identification, *International Journal of Control* 56 (1992) 319–346.
[10] Y.A. He, X.Z. Hu, S. Zhan, Predicting seismic response of structures by artificial neural networks, *Transactions of Tianjin University* 2 (1996) 38–41.
[11] G. Lightbody, W. Irwin, Multi-layer perceptron based modelling of nonlinear systems, *Fuzzy Sets and Systems* 79 (1996) 93–112.
[12] S. Jagannathant, F.L. Lewis, Identification of nonlinear dynamical systems using multilayered neural networks, *Automatica* 32 (1996) 1707–1712.
[13] J.S. Pei, A.W. Smyth, E.B. Kosmatopoulos, Analysis and modification of Volterra/Wiener neural networks for the adaptive identification of non-linear hysteretic dynamic systems, *Journal of Sound and Vibration* 275 (2004) 693–718.
[14] A.W. Smyth, S.F. Masri, E.B. Kosmatopoulos, A.G. Chassiakos, T.K. Caughey, Development of adaptive modeling techniques for nonlinear hysteretic systems, *International Journal of Non-Linear Mechanics* 37 (2002) 1435–1451.
[15] A. Yazdizadeh, K. Khorasani, Adaptive time delay neural network structures for nonlinear system identification, *Neurocomputing* 47 (2002) 207–240.
[16] C.S. Huang, S.L. Hung, C.M. Wen, T.T. Tu, A neural network approach for structural identification and diagnosis of a building from seismic response data, *Earthquake Engineering and Structural Dynamics* 32 (2003) 187–206.
[17] B. Xu, Z.S. Wu, G.D. Chen, K. Yokoyamab, Direct identification of structural parameters from dynamic responses with neural networks, *Engineering Applications of Artificial Intelligence* 17 (2004) 931–943.
[18] V.M. Becerra, F.R. Garces, S.J. Nasuto, W. Holderbaum, An efficient parameterization of dynamic neural networks for nonlinear system identification, *IEEE Transactions on Neural Networks* 16 (2005) 983–988.
[19] Y.C. Liang, D.P. Feng, J.E. Cooper, Identification of restoring forces in non-linear vibration systems using fuzzy adaptive neural networks, *Journal of Sound and Vibration* 242 (2001) 47–58.
[20] R. Le Riche, D. Gualandris, J.J. Thomas, F.M. Hemez, Neural identification of non-linear dynamic structures, *Journal of Sound and Vibration* 248 (2001) 247–265.
[21] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, 1995.
[22] K. Worden, A.J. Lane, Damage identification using support vector machines, *Smart Materials and Structures* 10 (2001) 540–547.
[23] K. Worden, G. Manson, The application of machine learning to structural health monitoring, *Philosophical Transactions of the Royal Society of London Series A* 365 (2007) 515–537.
[24] J. Zhang, T. Sato, S. Iai, Novel support vector regression for structural system identification, *Structural Control and Health Monitoring* 14 (2007) 609–626.
[25] H.S. Tang, S.T. Xue, R. Chen, T. Sato, Online weighted LS-SVM for hysteretic structural system identification, *Engineering Structures* 28 (2006) 1728–1735.
[26] J. Zhang, T. Sato, Non-linear hysteretic structural identification by utilizing on-line support vector regression, *Doboku Gakkai Ronbunshuu A* 62 (2006) 312–322.
[27] A.J. Smola, Learning with Kernels, PhD Thesis, Technical University Berlin, 1998.
[28] L. Ljung, *System Identification: Theory for the User*, second ed., Prentice-Hall, New York, 1999.
[29] S.M. Pandit, *Modal and Spectrum Analysis: Data Dependent Systems in State Space*, Wiley, New York, 1991.
[30] P. Andersen, R. Brincker, P. Kirkegaard, Theory of covariance equivalent ARMAV models of civil engineering structures, *Proceedings of IMAC 14*, Michigan, 1996, pp. 518–524.
[31] P. Andersen, Identification of Civil Engineering Structures Using ARMA Models, PhD Thesis, Aalborg University, 1997.
[32] Y.F. Dong, Y.M. Li, Application of time series analysis on performance evaluation of seismic structures, *Proceedings of the Ninth International Symposium on Structural Engineering for Young Experts*, Vol. 1, 2006, pp. 820–826.
[33] http://www.civil.eng.buffalo.edu/idarc2d50/introduction.htm.