TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.7, July 2014, pp. 5235 ~ 5243
DOI: 10.11591/telkomnika.v12i7.5787

Wavelet Kernel Based on Identification for Nonlinear Hybrid Systems

Hamid Nourisola
Department of Electrical control Engineering, Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, IRAN
email: hamidnourisola@yahoo.com

Abstract

This paper presents a new method based on wavelet for a class of nonlinear hybrid systems identification. Hybrid systems identification is composed of two problems; estimate the discrete modes or switch among the system modes and estimate continues submodels. In this paper, we assumed that haven't any prior knowledge about data classification and submodels identification. Also the combining of feature vector selection algorithm and wavelet are used in subspace learning and support vector machine as a classifier. The results indicate that the error of using the wavelet in subspace learning process becomes low. In addition, the proposed method is convergent and has an acceptable response in presence of high-power noise.

Keywords*: hybrid system identification, wavelet kernel function, feature vector selection, support vector machine classifier*

1. Introduction

Hybrid systems switch among several continues modes described as systems which includeboth continue and discrete states. In many application, an accurate model of system is not available, thus it is necessary to identify system parameters and their dynamics. In this paper, a class of nonlinear hybrid system identification in nonlinear autoregressive with external input (NARX) form is considered as follows:

$$y_i = f_{\lambda_i}(x_i) + e_i \tag{1}$$

Where e_i is an additive Gaussian noise term and $x_i = [y_{i-1},...,y_{i-n_a},u_{i-n_k},...,u_{i-n_k-n_c+1}]^T$ is continues state regression. n_c, n_a are lagged in outputs y_{i-k} and inputs y_{i-n_k-k} respectively. The discrete modes are determined by $\lambda_i \in \{1,2,...,n\}$ in which the one of n submodels $\{f_j\}_{j=1}^n$ is active at time step i. Also the number of modes is known and any information about their regressors is not available.

In [1], five methods of hybrid systems identification has been studied which these methods are non-convergent; in addition, optimization problem is enormously dependent on the initial condition. The mixed-integer programming is one of the mentioned methods which responses are limited to the number of data and variables [2, 3]. In [4], the number of modes is known and proposes an identification algorithm which combines clustering, regression and classification techniques. Unknown parameters of Bayesian approach are considered as random variables presented in [5]. This method has a three step: parameter estimation, data classification, and estimates of region and Bayesian law are inferred to estimate the parameters. In Algebraic Geometric approach, the applied system is assumed without noise [6]. The mentioned approach has obviously considerable error under the experimental systems. Bounded-error approach identifies the hybrid systems through imposing the error constriction [7]. In [8], formula construction is used as a least square problem with sum-of-norms regularization over regressor parameter differences. Automatic tuning approach applies

Received February 13, 2014; Revised March 16, 2014; Accepted April 2, 2014

bounded-error approach and support vector regression (SVR) for extension of the algebraic method [9]. [10] uses algebraic and SVR approaches to establish a framework based on minimizing the product of loss functions along with a regularization term. Lauer is presented another method for hybrid systems identification based on support vector classifier and kernel function [11, 12]. Kernel function is used as a nonlinear transformation. In [13], Luange proposes four methods for feature extraction and uses them in SVM formula to identify the nonlinear hybrid systems. In addition, each mode has a different radial basic function covariance in which modes, train and test data are known in subspace mapping step. In [14] a new learning approach for piece wise smooth functions by regularized kernel regression is proposed. This is done by defining a new regularization term. In [15] identification of hybrid systems involving arbitrary and unknown nonlinearities in the submodels is investigated. In this approach, the submodels are estimated one by one by maximizing the sparsely of the corresponding error vector.

In this paper, the proposed method improves the work of [13] with wavelet function. The main contribution of this paper is the change in the subspace learning form train and test data using the wavelet kernel function in nonlinear hybrid systems identification. Finally, the effect of kernel function coefficient and wavelet kernel function coefficient are investigated. In this paper, we assume that train and test data are unknown in data calssification and data have single RBF covariance. We know only about the number of modes.

This paper is organized as follows: In Section 2, a framework of nonlinear hybrid systems identification is introduced. Section 3 presents the kernel principal component regression and wavelet kernel principal component regression. In Section 4, proposed method with numerical results is investigated. Finally, the conclusions are drawn in section 5.

2. Framework of Nonlinear Hybrid Systems Identification

First, This section presents the structure of kernel function for submodels estimating [12, 13]. Nonlinear hybrid systems submodels can introduce as:

$$f_j(x) = \sum_{k=1}^{N} \alpha_{kj} k_j(x_k, x) + b_j \quad (2)$$

Where α_j includes $[\alpha_{1j},...,\alpha_{Nj}]^T$, b_j is bias term for f_j and k_j is kernel function that satisfy the mercer conditions [16]. Typical kernel functions are linear kernel function, RBF kernel function and polynomial kernel function. In this paper, RBF kernel function $k(x_k, x) = \exp(-\|x_k - x\|_2^2 / 2\delta^2)$ is used.

The method mentioned in [17-19] for identification and data classification is (3).

$$R_{reg}(w) = T(w) + cR_{emp}(w) \quad (3)$$

Where $R_{emp}(w)$ is empirical risk function. $T(w)$ is a term which prevents the extra training and during minimizing the empirical risk function acts adjustment perform. c is the adjustment coefficient. According to (4), the aim of empirical risk function is minimizing the number and limit of classification error.

$$R_{emp}(w) = \frac{1}{N} \sum_{i=1}^{N} q(y_i - f(x, w)) \quad (4)$$

Where y_i, $f(x, w)$ and q are class labels, output of classifier and weighted function respectively in support vector machines classification. According to (4), (3) is rewritten as follows:

$$w_{primal} = \min \frac{1}{2}\|w\|^2 - c\sum_{i=1}^{N} \alpha_i ((w^T x_i + b) y_i - 1) \quad (5)$$

Where $w = \sum_{i=1}^{N} \alpha_i y_i x_i$ and $\|.\|$ is norm 1 or 2. To obtain the solution, the dual optimization problem must be solved with boundary expressed conditions by:

$$\begin{aligned} w_{dual} &= \max \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j x_i^T x_j \\ s.t. \quad & 0 \le \alpha_i \le c \ , \ 1 \le i \le N \\ & \sum_{i=1}^{N} \alpha_i y_i = 0 \end{aligned} \quad (6)$$

In (6), the input data can be in another space. This means that data have been mapped to another space. When the data have nonlinear behavior and cannot distinguish among them, data mapping is used. Mode estimation for each data obtains through (7).

$$\lambda_i = \arg\max_i \ f_i(x) \quad (7)$$

Where $f_i(x) = \sum_{i=1}^{N} \alpha_i y_i x_i^T x + b_i$.

3. Subspace Learning and Data Dimension Reduction

The purpose of this section is to reduce the number of added data which describe the system with extra features. Subspace learning is used through selecting the eigenvalues and eigenvectors of the training data matrix and evaluating its effect on test data. Then use the Support Vector Machine Lagrangian multipliers $\{\alpha_{ij}\}_{j=1,...,n}$ for each test vectors in each mode obtained. One of the support vector machine properties is reducing the number of test data. In fact, the number of data in two steps (subspace learning and classification) decreases. The operations of two steps are explained in the following.

3.1. Kernel Principal Component Regression

If the data distribution has nonlinear behavior in the original space, it cannot be changed by linear mapping. So it is necessary to use nonlinear mapping to reduce the nonlinear relation between the data.

As mentioned in [13], [20-21], Kernel Principal Component Regression (KPCR) can reduce the training data dimension in optimal way. Suppose that a set of training feature vectors in the original space is $\{z_1, z_2, ..., z_N\}$ where $z_i \in R^n \ (1 \le i \le N)$ is feature extraction from mode i. Also suppose that $\phi : R^n \to g$ is a nonlinear transformation which transforms data from the original space of dimension n to a feature space of dimension l. In this space, scatter matrix is obtained according to (8).

$$S_t = \frac{1}{N}\sum_{i=1}^{N} (\varphi(z_i) - e)(\varphi(z_i) - e)^T \quad (8)$$

Where $\varphi(z_i) \in f \ (1 \le i \le N)$ is vector i in feature space and e is the average of all vectors in the feature space. If the mean of vectors is not zero, we can transfer kernel function in feature space to zero through (9).

$$\hat{k}_j = (I_N - \frac{1}{N}1_{N\times N})k_j(I_N - \frac{1}{N}1_{N\times N}) \quad (9)$$

Where I_N and $1_{N\times N}$ are Identity and unit matrix respectively.

Eigenvalues and eigenvectors matrix of S_t placed out of null space g can be calculated through using the PCR algorithm.

$$S_t v = \lambda v \quad (10)$$

Where $v = \sum_{i=1}^{N} \alpha_i \varphi(z_i)$. According to [21], (10) will be as (11).

$$kW = N\lambda W \quad (11)$$

Where $W = [\alpha_1, \alpha_2, ..., \alpha_N]^T$ and k is kernel function matrix. If above operation (11) for m large eigenvalues of S_t is done, m vectors $w_1, w_2, ..., w_m$ will be obtained. It is obvious that to obtain the m vectors, we must compute the eigenvalues and eigenvectors of kernel matrix k. Finally, mapping of each test feature vectors such as $z \in R^n$ from original space with n dimension to m-dimensional subspace ($m < n$) is done by (12).

$$x = W^T k_z \quad (12)$$

Where $x \in R^m$ is the mapping of test feature vector in subspace and $k_z = [k(z, z_1), k(z, z_2), ..., k(z, z_N)]^T$.

To obtain the kernel principal component regression and identification of system modes:

1. Compute the kernel matrix for a set of training data.
2. Compute eigenvalues and eigenvectors for kernel matrix determine its dimension by (13) and calculate the transfer matrix.

$$\frac{\sum_{i=1}^{m} s_i}{\sum_{i=1}^{N} s_i} \geq \tau \quad (13)$$

Where m is matrix dimension, s_i is eigenvalue of kernel matrix and $\tau \in [0,1]$ relates system error to its dimension. The amount of this parameter can be decreased until the system error is low.

3. Transfer the test data to feature space using the mapping matrix.
4. Place the transferred test data in (6).
5. Calculate $\hat{f}$ for each data set using (14) and then identify and classify them by (6) and (7).

$$\hat{f}_j(x) = \alpha_j^T W_j^T k_j(., x) + b_j \quad (14)$$

Where α_j is lagrangian coefficient in feature space.

3.2. Wavelet Kernel Principal Component Regression

The wavelet kernel principal component regression is the extension of KPCR method. This method can be used for nonlinear systems. In this paper, wavelet transform is used for

subspace learning. The main idea of the wavelet analysis is to approximate functions by dilations and translations function $h(x)$ called the mother wavelet.

$$h_{a,c}(x) = \left|x\right|^{-1/2} h(\frac{x-c}{a}) \tag{15}$$

Where $x, a, c \in R$. a is a dilation factor and c is a translation factor. If the wavelet function is multidimensional, it can rewrite as (16) [22].

$$h(x) = \prod_{i=1}^{N} h(x_i) \tag{16}$$

Where $x = (x_1, x_2, ..., x_N) \in R^N$. If $x, x' \in R^N$ are two vectors in a space, wavelet kernel can be obtained from (17) [23].

$$k(x, x') = \prod_{i=1}^{N} h(\frac{x_i - c_i}{a}) h(\frac{x_i' - c_i}{a}) \tag{17}$$

In translation-invariant kernel $k(x, x') = k(x - x')$, (17) can be rewritten as (18).

$$k(x, x') = \prod_{i=1}^{N} h(\frac{x_i - x_i'}{a}) \tag{18}$$

Mother wavelet function is assumed the form $h(x) = [\cos(1.75x)\exp(-\frac{x^2}{2})]^P$, kernel function is as (19).

$$\begin{aligned} k(x, x') &= \prod_{i=1}^{N} [h(\frac{x_i - x_i'}{a})]^P \\ &= \prod_{i=1}^{N} [\cos(1.75(\frac{x_i - x_i'}{a}))\exp(-\frac{\left\|x_i - x_i'\right\|^2}{2a^2})]^P \end{aligned} \tag{19}$$

Wavelet kernel is an orthonormal function [24] while this feature is not in Gaussian kernel function. In other words, due to the dependencies and correlations between data in the Gaussian kernel function, train speed will be lower than the wavelet kernel.

WKPCR algorithm is applied in the same way as well as the KPCR algorithm except that their kernel function will be different. In feature extraction and subspace learning for nonlinear systems, wavelet kernel function is used. The process of the algorithm is as follows:

1. Compute the wavelet kernel matrix for a set of training data.
2. Check mean of the train data in mapping subspace using $\sum_{i=1}^{N} \psi_j(x_i)$ where $\psi_j(x_i)$ is i[th] column of mapping train data matrix. If the mean of train data is not zero, the wavelet kernel function must be transferred by (9).
3. Compute the mapping matrix using the eigenvectors of wavelet kernel matrix.
4. Transfer the test data and place them in (6).
5. Calculate $\hat{f}$ for each data set using (14) and then identify and classify them by (6) and (7).

4. Simulation Results

This section involves the estimation of a function which switches among four unknown nonlinear systems. Consider the function arbitrarily switches among four nonlinear behaviors as (20). Estimation of the system is given in Figure 1.

$$f(x)=\begin{cases} x^2 & \lambda_1 \\ \sin(3x)+2 & \lambda_2 \\ \cos(2x)+3 & \lambda_3 \\ \sqrt{x^3+1} & \lambda_4 \end{cases} \quad (20)$$

The results of the KPCR [12, 13] and WKPCR for different values of their parameters are expressed in tables 1 and 2.

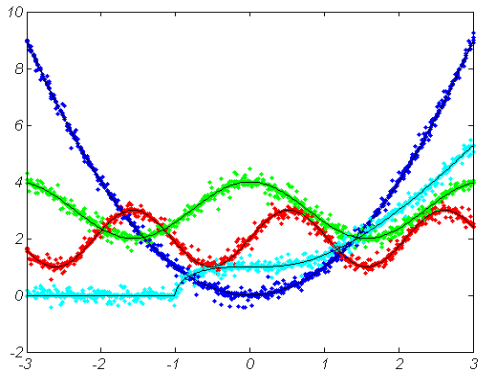

Figure 1. Estimation of a Switched Nonlinear Function from 2000 Noisy Data Points

A training set of N = 2000 points is generated by (20) with additive zero-mean Gaussian noise (standard deviation $\sigma \in [0, 0.7]$) for uniformly distributed random $x \in [-3, 3]$ and uniformly distributed random $\lambda_i \in \{1, 2, 3, 4\}$. The number of train and test data is 100 and 300 respectively. This system is identified by KPCR method and its results are given in Table 1. Data classification error obtained from multi-class support vector machines classification method and confused matrix shown in Table 1. Confused matrix shows the data which have been classified wrongly are place in which modes; the share of data classified incorrectly becomes high as the similarity of the shape and type of modes increases. This case occurs between modes 2 and 3.

Table 1. Result of KPCR Method on Switched Nonlinear System

δ	*Test. Classif. Error %*	*Confused Matrix*
5×10^2	2.1 ± 0.35	$\begin{bmatrix} 147.9\pm1.05 & 1.5\pm0.8 & 0.6\pm0.3 & 0 \\ 0 & 148.3\pm0.8 & 1.5\pm0.5 & 0 \\ 0 & 8.8\pm1.25 & 141.1\pm1.2 & 0 \\ 0 & 0.2\pm0.3 & 0.4\pm0.35 & 148.9\pm0.5 \end{bmatrix}$
10^3	2.45 ± 0.45	$\begin{bmatrix} 147.3\pm1.6 & 2.05\pm1.25 & 0.6\pm0.6 & 0 \\ 0.2\pm0.35 & 147.9\pm1.1 & 2\pm0.6 & 0 \\ 1.2\pm0.8 & 0.9\pm0.25 & 148.2\pm2.65 & 0 \\ 0 & 0.2\pm0.25 & .5\pm0.3 & 149.3\pm0.5 \end{bmatrix}$
10^4	1.28 ± 0.49	$\begin{bmatrix} 150 & 0 & 0 & 0 \\ 0.4\pm0.7 & 145.3\pm3.2 & 4.3\pm3.1 & 0.2\pm0.63 \\ 0.4\pm0.45 & 2.6\pm1.8 & 147\pm1.8 & 0 \\ 0.1\pm0.32 & 0 & 0 & 149.2\pm0.2 \end{bmatrix}$

Table 2. Result of WKPCR Method on Switched Nonlinear System

α	***Test. Classif. Error %***	***Confused Matrix***
10^2	12.88 ± 0.88	$\begin{bmatrix} 131.9 \pm 1.37 & 1.4 \pm 1.26 & 0.3 \pm 0.67 & 16.4 \pm 1.51 \\ 15 \pm 2.62 & 131.6 \pm 2.84 & 0.7 \pm 0.82 & 2.7 \pm 1.34 \\ 11.4 \pm 2.59 & 3.5 \pm 1.65 & 127.1 \pm 2.23 & 8 \pm 2.36 \\ 18.2 \pm 1.4 & 0.5 \pm 0.71 & 0.7 \pm 0.95 & 130.6 \pm 1.9 \end{bmatrix}$
2×10^2	2.27 ± 0.65	$\begin{bmatrix} 146.3 \pm 1.89 & 1.5 \pm 0.85 & 2.2 \pm 1.55 & 0 \\ 0 & 148.2 \pm 1.32 & 1.6 \pm 1.08 & 0 \\ 0 & 4.3 \pm 1.57 & 145.7 \pm 1.57 & 0 \\ 0 & 3.4 \pm 2.12 & 0.4 \pm 0.7 & 146.8 \pm 1.87 \end{bmatrix}$
5×10^2	3.03 ± 0.64	$\begin{bmatrix} 143.7 \pm 1.95 & 2.3 \pm 1.5 & 3.8 \pm 2.53 & 0 \\ 0 & 147.8 \pm 1.14 & 2.2 \pm 1.14 & 0 \\ 0 & 4.3 \pm 1.16 & 145.7 \pm 1.16 & 0 \\ 0 & 5.1 \pm 1.91 & 0.9 \pm 0.74 & 144 \pm 1.89 \end{bmatrix}$
10^4	1.17 ± 0.32	$\begin{bmatrix} 150 & 0 & 0 & 0 \\ 0 & 146.67 \pm 3.2 & 3 \pm 3.46 & 0.33 \pm 0.82 \\ 0 & 3.17 \pm 2.86 & 147.17 \pm 3.12 & 0 \\ 0 & 0.5 \pm 1.22 & 0 & 149.5 \pm 1.12 \end{bmatrix}$

The data of Table 1 is shown that the data classification error decreases with the increment of δ. This error is reduced to a specified value of δ, then the error will be increased. Actually, the system error has a minimum point in RBF parameter.

In WKPCR method, the number of train and test data are 50 and 150, the value of c and P are 1000 and 1 respectively. The data classification error and confused matrix results are shown in Table 2. The classification error for small amount of α is high and classification error decreases considerably as increase of α. This method has a minimum point for classification error in wavelet parameter. If the value of p is not fixed, for a fixed value of wavelet kernel parameter such as $\alpha = 10^3$, error changes for different values of P is shown in Figure 2.

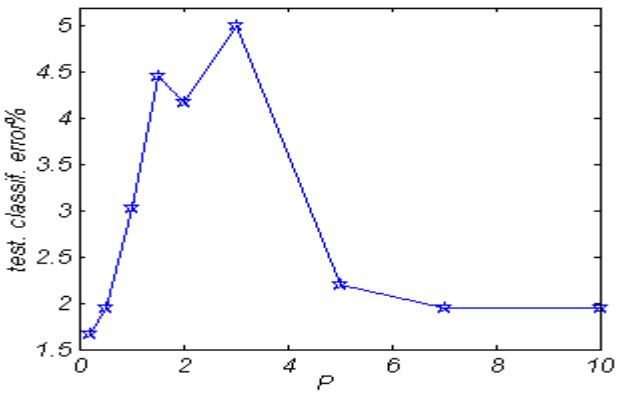

Figure 2. Error Changes for Different Values of P

The classification error for two methods is low as 15% as shown in Table 1 and 2. When the wavelet is used, the classification error is low in comparison with the case that the wavelet isn't applied. Figure 3 shows the classification error in the same conditions for two methods.

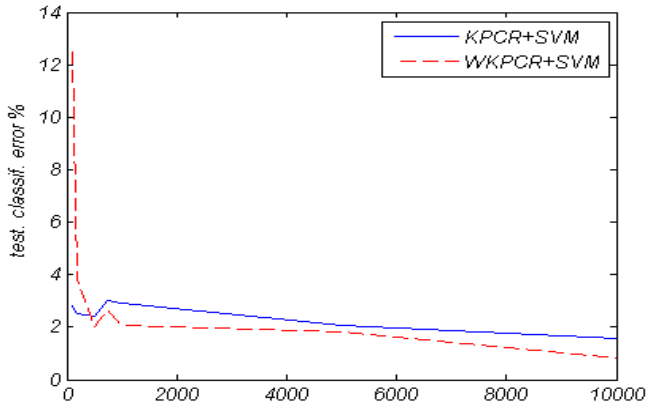

Figure 3. Comparison of KPCR [12, 13] and WKPCR Classification Error in the Same Conditions

5. Conclusion

In this paper, a new method based on wavelet for identification and submodel estimation of nonlinear hybrid systems is proposed. Selected wavelet kernel function is multidimensional. This method could approximate a nonlinear hybrid system and can be implemented on the hybrid systems which switch among unknown modes. Estimating the number of submodels and data classification for linear and nonlinear hybrid systems is important issue presented in this paper. Dependence among kernel function training data causes reduction of learning speed. Furthermore, proposed method eliminates this dependency and improves learning speed.

Further investigation will focus on optimizing and selecting the regular parameter c, choosing an appropriate kernel function for subspace learning and using the nonlinear support vector classification for mixed data.

References

[1] Paoletti S, Juloski A, Ferrari-Trecate G, Vidal R. Identification of hybrid systems: A tutorial. *Euopean Journal of Control.* 2007; 13(2–3): 242–262.
[2] Bemporad A, Roll J, Ljung L. *Identification of hybrid systems via mixed-integer programming. Proceedings* of the 40th IEEE Conference on Decision and Control, 2001; 1: 786-792.
[3] Roll J, Bemporad A, Ljung L. Identification of piecewise affine systems via mixed-integer programming. *Automatica*. 2004; 40(1): 37-50.
[4] Ferrari-Trecate G, Muselli M, Liberati D. Morari M. A clustering technique for the identification of piecewise affine systems. *Automatica.* 2003; 39(2): 205-217.
[5] Juloski AL, Weiland S, Heemels WPMH. A Bayesian approach to identification of hybrid systems. *IEEE Transactions on Automatic Control.* 2005; 50(10): 1520-1533.
[6] Vidal R, Soatto S, Ma Y, Sastry S. *An algebraic geometric approach to the identification of a class of linear hybrid systems.* 42nd IEEE Conference on Decision and Control Proceedings. Berkeley. 2003; 1: 167-172.
[7] Bemporad A, Garulli A, Paoletti S, Vicino A. A bounded-error approach to piecewise affine system identification. *IEEE Transactions on Automatic Control*. 2005; 50(10): 1567-1580.
[8] Ohlsson H, Ljung L, Ohlsson H. *Identification of Piecewise Affine Systems Using Sum-of-Norms Regularization.* 18th IFAC World Congress. Milano. 2011; 6640-6645.
[9] Lauer F, Bloch G. *A new hybrid system identification algorithm with automatic tuning.* 17th IFAC World Congress. Seoul. 2008; 10207-10212.
[10] Lauer F, Vidal R, Bloch G. *A product-of-errors framework for linear hybrid system identification*. Proceedings of the 15th IFAC symp. on system identification. Saint-Malo. 2009.
[11] Lauer F, Bloch G. Switched and piecewise nonlinear hybrid system identification In Hybrid Systems: Computation and Control. *Springer Berlin Heidelberg.* 2008; 4981: 330-343.
[12] Lauer F, Bloch G, Vidal R. *Nonlinear hybrid system identification with kernel models.* 49th IEEE Conference on Decision and Control (CDC). Atlanta. 2010; 696-701

[13] Luong Le V, Bloch G, Lauer F. Reduced-size kernel models for nonlinear hybrid system identification *IEEE Transactions on Neural Networks.* 2011; 22(12): 2398-2405.
[14] Lauer F, Le VL, Bloch G. *Learning smooth models of nonsmooth functions via convex optimization*. IEEE International Workshop on Machine Learning for Signal Processing (MLSP). Santander. 2012; 1-6.
[15] Le VL, Lauer F, Bako L, Bloch G. *Learning nonlinear hybrid systems: from sparse optimization to support vector regression.* Proceedings of the 16th international conference on Hybrid systems: computation and control (ACM). Philadelphia. 2013; 33-42
[16] Vapnik Vladimir N. An overview of statistical learning theory. *IEEE Transactions on Neural Networks.* 1999; 10(5): 988-999
[17] Meshgini s. Automatic Face Recognition using Support Vector Machines Ph.D. Thesis. University of Tabriz. 2013.
[18] Abe S. Support vector machines for pattern classification. *Springer*. 2010.
[19] Burges CJ. A tutorial on support vector machines for pattern recognition*. Date mining and knowledge discovery 2.* 1998; 2: 121-167.
[20] Scholkopf B, Smola A, Muller KR. Nonlinear component analysis as kernel eigenvalue problem. Neural Computation. 1998; 10(5): 1299-1319.
[21] Rosipal R, Girolami M, Trejo L, Cichocki A. Kernel PCAfor feature extraction and de-noising in non-linear regression. *Neural Computing & Applications*. 2001; 10(3): 231–243.
[22] Zhang QH, Benveniste A. Wavelet networks. IEEE Transaction on Neural Networks. 1992; 3: 889-898.
[23] Zhang L, Zhou W, Jiao W. Wavelet support vector machine. IEEE Transaction on System Man and Cybernetics. 2004; 34(1).
[24] Aubechies I. Orthonormal bases of compactly supported wavelets. *Communications on pure and applied mathematics.* 1988; 41(7): 909-996.