Autoregressive Modeling and PCA Preprocessing to Support Vector Machines Based on PSO for Bearing Fault Diagnosis

T. Thelaidjia
Welding and NDT Research Center
Cheraga, Algeria
t.thelaidjia@csc.dz

A. Moussaoui
University of 08 Mai 1985
Guelma, Algeria

S. Cheneikher
University of Tebessa
Tebessa, Algeria

A. Boutaghane
Welding and NDT Research Center
Cheraga, Algeria
a.boutaghane@csc.dz

***Abstract* - In this paper a method for fault diagnosis of rolling bearings is presented. It consists of two major parts: vibration signal feature extraction and condition classification for the extracted features. In this paper Autoregressive Modeling followed by Principal Components Analysis (PCA) was introduced for feature extraction from faulty bearing vibration signals. After extracting feature vectors by AR-PCA, the support vector machine (SVM) was applied to automate the fault diagnosis procedure. To improve the classification accuracy for bearing fault prediction, particle swarm optimization (PSO) is employed to simultaneously optimize the SVM kernel function parameter and the penalty parameter. The results have shown feasibility and effectiveness of the proposed approach.**

Index Terms - Autoregressive Modeling, Principal Components Analysis, Support Vector Machine, Particle Swarm Optimization, Wavelet Packet, Fault Diagnosis, Roller Bearing.

I. INTRODUCTION

Bearings are frequently applied components in the vast majority of rotating machines. Their running quality influences the working performance of equipment. Statistically, 30% of rotational mechanical equipments malfunction is caused by the faults in bearings [1]. Therefore, many important researches had been done in the advanced field of bearing fault diagnosis [1], [2], [3], [4]. Using the vibration signals of rolling bearings and components to monitor and diagnose their working state, is the common used method in the study of bearing fault diagnosis [1], [3]. Support Vector Machine (SVM) is a new machine learning method which was introduced by Vapnik on the foundation of statistical learning theory (SLT). However, since the middle of 1990s, the algorithms used for SVM started emerging with greater availability of computing power [5], [6]. The main difference between the known domain of artificial neural network (ANN) and SVM is in the principle of risk minimization (RM) [2], [3]. In case of SVM, structural risk minimization (SRM) principle is used to minimize an upper bound on the expected risk whereas in ANN, traditional empirical risk minimization (ERM) is used to minimize the error on the training data. The difference in RM leads to better generalization performance for SVM than ANN. According to the literature, SVM has been successfully applied to many applications, such as pattern identification, regression analysis, function approximating, etc [7], [8], [9], [10]. The results give the evidence that the technique is not only quite satisfying from a theoretical point of view, but also can lead to high performance in practical applications.

Finding out good features is an important phase in distinguishing the different mechanical failure; As an interesting example, Wavelet Packet analysis has been utilized for impulse mechanical failure classification [1]. Parameter optimization is the key to perform SVM. At present, the widely used methods of parameter optimization for SVM are network search method, K-order cross-validation method, Leave-one-out method, etc. These algorithms have the disadvantage of huge amount of computation, and the calculated parameters are not always the best. In recent years, a series of intelligent bionic algorithms are proposed based on the biological behavior study in the natural, such as genetic algorithm (GA) and particle swarm optimization (PSO) [11], [12], [13].

PSO was proposed by Kennedy and Eberhart [14], [15]. And it is inspired by the social behavior of bird flocking, fish schooling and swarm theory, etc. The theoretical framework of PSO is very simple, and PSO possesses the properties of easy implementation and fast convergence [13], [16].

In this paper, the higher dimension time series data is extracted by autoregressive modeling then the state eigenvectors are reduced by principal components analysis (PCA), finally the resulting vectors are being used as an input of a multi fault classifier which is composed of SVM. Because particle swarm optimization is powerful, easy to implement, and computationally efficient [15], this study introduces PSO as an optimization technique to simultaneously optimize the SVM parameters.

II. FEATURE EXTRACTION

The fault diagnosis is essentially a problem of pattern recognition, of which, an important step is feature extraction.

In this study, three types of feature extraction are applied :Wavelet packet transform, Autoregressive modeling and PCA.

A. WAVELET PACKET ALGORITHM

The step of feature extraction based on three layer wavelet packet is given as follows :

Firstly,The vibration signal x(t) was decomposed by a mother wavelet, the signal features in eight frequency bands from low to high were extracted in the third layer. Secondly, The signal in each frequency band is extracted and the wavelet packet decomposition coefficient was reconstructed. D_{30} presents the reconstructed signal of d_{30}, D_{31} presents the reconstructed signal of d_{31}, and so on. The composed signal is defined as :

$$D = D_{30} + D_{31} + D_{32} + D_{33} + D_{34} + D_{35} + D_{36} + D_{37} \quad (1)$$

Finally, The signal energy of each frequency band is calculated as :

$$E_{3j} = \sum_{k=1}^{n} \left| d_{jk} \right|^2 \quad (2)$$

Normalized, Let : $T = \sum_{j=1}^{7} E_{3j}$

Eigenvector E was constructed based on each frequency band energy:

$$E = [\frac{E_{30}}{T}, \frac{E_{31}}{T}, \frac{E_{33}}{T}, \frac{E_{34}}{T}, \frac{E_{35}}{T}, \frac{E_{36}}{T}, \frac{E_{37}}{T} \quad (3)$$

B. LEAST-SQUARE METHOD FOR AR PARAMETER ESTIMATION.

In this section, we derive a method of AR estimator, witch based on a least-squares (LS) minimization criterion using the time-domain relation $A(z)y(t) = e(t)$ [17], [18]. Let $x(n)$ be an AR process of order p. Then $x(n)$ satisfies:

$$e(n) = x(n) + \sum_{k=1}^{p} \alpha_k x(n-k) = x(n) + \hat{x}(n) \quad (4)$$

We interpret $\hat{x}(n)$ as a linear prediction of $x(n)$. from the n previous samples $x(n-1), \dots x(n-p)$, and we interpret $e(n)$ as the corresponding prediction error.

The vector $\alpha = [\alpha_1, \dots, \alpha_p]'$ that minimizes the prediction error $\rho = E\{|e(n)|^2\}$ is the AR coefficient vector,we have:

$$\begin{aligned} \rho &= E\{|e(n)|^2\} = E\{|\hat{x}(n) - x(n)|^2\} \\ &= r_{xx}(0) + \alpha^H r + r^H \alpha + \alpha^H R \alpha \end{aligned} \quad (5)$$

where α, R, r are defined by:

$$\alpha = [\alpha_1, \dots, \alpha_p]' \quad (6)$$

$$r = [r_{xx}(1), \dots, r_{xx}(p)] \quad (7)$$

$$R = \begin{bmatrix} r_{xx}(0) & r_{xx}(-1) & \cdots & r_{xx}(-p+1) \\ r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(-p+2) \\ \vdots & \vdots & \vdots & \vdots \\ r_{xx}(p-1) & r_{xx}(p-2) & \cdots & r_{xx}(0) \end{bmatrix} \quad (8)$$

The vector α that minimizes (5) is given by:

$$\alpha = -R^{-1} r \quad (9)$$

with corresponding minimum prediction error :

$$\rho = r_{xx}(0) - r^H R^{-1} r \quad (10)$$

The least-squares AR estimation method is based on a finite sample approximate solution of the above minimization problem. Given a finite set of measurements $\{x(n)\}_{n=1}^{N}$ we approximate the minimization $\rho = E\{|e(n)|^2\}$ by the finite sample cost function:

$$f(\alpha) = \sum_{n=n_0}^{n_1} |e(n)|^2$$

$$f(\alpha) = \sum_{n=n_0}^{n_1} \left| x(n) + \sum_{k=1}^{p} \alpha_k x(n-k) \right|^2 \quad (10)$$

$$f(\alpha) = \| h + X\alpha \|^2 \quad (12)$$

Such that:

$$h = \begin{bmatrix} x(n_0) \\ x(n_0+1) \\ \vdots \\ x(n_1) \end{bmatrix};$$

$$X = \begin{bmatrix} x(n_0-1) & \cdots & x(n_0-p) \\ x(n_0) & \cdots & x(n_0+1-p) \\ \vdots & \vdots & \vdots \\ x(n_1-1) & \cdots & x(n_1+1-p) \end{bmatrix};$$

where we assume x(n) = 0 for n < 1 and n > N The vector α that minimizes f(α) is given by:

$$\alpha = -(X * X)^{-1}(X * h \quad (13)$$

where, as seen from (12) the definitions of X and h depend on the choice of (n_0, n_1), when $n_0 = p+1$ and $n_1 = N$ this choice is often named the covariance method.

C. PRINCIPAL COMPONENTS ANALYSIS (PCA) [3]

The time series of vibration signal can be written as $x_i (i = 1,2,\dots,n)$, m linear independent variables $y_i (i = 1,2,\dots,m)$, $m \le n$ can be obtained by SVD to denote x_i, then:

$$X = BY \quad (14)$$

where B is $N * N$ transformation matrix.

With the matrix transformation, so we have:

$$E(XX^T) = BE(YY^T)B^T \quad (15)$$

$$R_x = BR_yB^T \quad (16)$$

Since autocorrelation matrix R_x is real symmetric matrix, autocorrelation matrix R_y can be changed to a diagonal matrix which is composed of n positive real eigenvalues $\lambda_i (i = 1, \dots, n)$ of R_x, then:

$$R_y = E[(\lambda_i)^2] = [\lambda_i] \quad (17)$$

Where $\lambda_1 > \lambda_2 > \cdots > \lambda_n$. We can choose m eigenvectors which corresponding the prior m eigenvalues, the m eigenvectors can constitute a m subspace.

In order to keep the m eigenvectors maintaining adequate original information, the m should be:

$$\sum_{i=1}^{m} \lambda_i / \sum_{i=1}^{n} \lambda_i > 85\% \quad (18)$$

It means that eigenvalues $\lambda_i (i = 1, \dots, m)$ compressed main information of the original signals, so we can achieve the fault pattern classified by the eigenvalues and eigenvectors research.

III. SUPPORT VECTOR MACHINE

The support vector machine (SVM) is a supervised learning method that generates input-output mapping functions from a set of labeled training data. For classification, nonlinear kernel functions are often used to transform input data to a high-dimensional feature space in which the input data become more separable compared to the original input space [19].

Support vector machine (SVM) based on statistical learning theory is proposed according to optimal hyper-plane in the case of linear separable [1].

If the hyper-plane separates all samples correctly, it must satisfy the following condition [4]:

$$y_k(\langle w; x\rangle - \lambda_0) \geq +1, \forall\, k \in \{1, \dots, n\} \quad (19)$$

In order to find the optimal hyper-plane, we need to minimize the following functional [4]:

$$\varphi(\omega) = \frac{1}{2}\|\omega\|^2 \quad (20)$$

Solution of the optimal problem is given by the saddles of Lagrange function as below:

$$L(\omega, \lambda_0, \alpha) = \frac{\|\omega\|^2}{2} - \sum_{k=1}^{n} \alpha_k [y_k(\langle w; x\rangle - \lambda_0) - 1] \quad (21)$$

where $\alpha = (\alpha_1, \dots, \alpha_n)$ is the Lagrange coefficient; $\alpha_i \geq 0, \forall$ i .

The original problem can be transferred to the dual problem as below:

$$W(\alpha) = \sum_{k=1}^{n} \alpha_k - \frac{1}{2}\sum_{k,k'=1}^{n} \alpha_k\, \alpha_{k'} y_k y_{k'} \langle x_k; x_{k'}\rangle \quad (22)$$

subject to:

$$\sum_{k=1}^{n} \alpha_k y_k = 0 \; et \; \alpha_k \geq 0$$

If α^* is the optimal solution, then:

$$\langle w^*; x\rangle = \sum_{k=1}^{n} \alpha_k^* y_k \langle x_k; x\rangle \quad (23)$$

It means that the weight coefficients of the optimal hyper-plane are the linear combination of the training sample vector. According to the Kuhn−Tucker condition, the solution of optimal problem must satisfy:

$$\alpha_*^k [y_k(\langle w^*; x\rangle - \lambda_0^*) - 1] = 0 \quad (24)$$

were λ_0^* is given by:

$$\lambda_0^* = \frac{1}{N_{sv}} \sum_{s=1}^{N_{sv}} (y_s - x_s^T \omega^*) \quad (25)$$

where N_{sv}: nomber of support vectors.

After solving the above problem, we can get the optimal classification function as below:

$$D(x) = sgn\left[\sum_{k \in N_{sv}} \alpha^* y_k x_i^T x - \lambda_0^*\right] \quad (26)$$

The nonseparable problem can be solved by soft-margin SVM [8], [20], [21].

If we used the inner $\kappa(x_k, x)$ substitute for the inner of the optimal hyperplane, the original feature space is mapped to new feature space [22]. And the optimal function can be formulated as below:

$$W(\alpha) = \sum_{k=1}^{n} \alpha_k - \frac{1}{2}\sum_{k,k'=1}^{n} \alpha_k\, \alpha_{k'} y_k y_{k'} \kappa(x_k, x_{k'})$$

subject to:

$$\sum_{k=1}^{n} \alpha_k y_k = 0 \; and \; 0 \leq \alpha_k \leq c$$

The corresponding decision function is written as below:

$$D(x) = sgn\left[\sum_{k \in N_{sv}} \alpha^* y_k \kappa(x_i^T x) - \lambda_0^*\right] \quad (27)$$

here, $\kappa(x_k, x)$ is called kernel function.

Usually, the kernel function can be expressed as below [7], [23].

Polynomial:

$$\kappa(x_1, x_2) = (1 + \langle x_1, x_2\rangle)^{\delta} \quad (28)$$

where "q" is the degree of the polynomial.

Radial basis function (RBF):

$$\kappa(x_1, x_2) = \exp(-\|x_1 - x_2\|^2 / 2\delta^2) \quad (29)$$

where δ^2 is the variance of the Gaussian function.

Sigmoid:

$$\kappa(x_1, x_2) = \tanh(\alpha_0 \langle x_1; x_2\rangle + \beta_0) \quad (30)$$

where: δ, α_0 and β_0 are the parameters of kernel function. The classification performances of SVM are affected by three techniques; the selecting of the kernel, the choosing of the kernel parameters, and the choosing of the regularization parameter "c" [4].

Most of cases in practical are multi-classed, such as in the rolling bearing classifying, it can be sorted into normal, outer race fault and inner race fault, etc. So, we have to design an approach to expend the application of SVM to a multi-classifying field because the SVM can deal with only two classes. The different combination principles constitute different classifying algorithm [7], [24], [25]. We employ the one-against-the-rest method to compose a multi-fault classifier. Since the SVM generalization performance heavily depends on the right setting of "c" and "σ", these two parameters need to be set properly by the user. According to the experience from numerical experiments [26], [27], "c" and "σ" exhibit a (strong) interaction. As a consequence, they should be optimized simultaneously, rather than separately.

IV. THE PARTICLE SWARM OPTIMIZATION

The particle swarm optimization (PSO) consists of a swarm of particles flying through the search space. Each particle is treated as a point in a D-dimensional space. The i^{th} particle is represented $Z_i = (Z_{i1}, Z_{i2}, \dots, Z_{id}, Z_{iD})$. The best previous position of any particle is recorded and represented as $P_i = (P_{i1}, P_{i2}, \dots, P_{id}, P_{iD})$. The index of the best particle among all the particles in the population is represented by the symbol G. The rate of the position change (velocity) for the i^{th} particle is represented as $V_i = (V_{i1}, V_{i2}, \dots, V_{id}, V_{iD})$. The updated velocity and position of the i^{th} particle at the k-th iteration are [11]:

$$V_{id}^{k} = w.V_{id}^{k-1} + c_1.r_1.\left(P_{id} - Z_{id}^{k-1}\right) + c_2.r_2.\left(P_{Gd} - Z_{id}^{k-1}\right) \quad (31)$$

$$Z_{id}^{k} = Z_{id}^{k-1} + V_{id}^{k} \quad (32)$$

Where c_1 and c_2 are constants knows as the cognitive and social acceleration coefficients, respectively, w is the inertia weight, r_1 and r_2 are random numbers between 0 and 1.

The first part of (31) represents the previous velocity, which provides the necessary momentum for particles to fly across the search space. The second part is the "cognition" part, which represents the private thinking of the particle itself.

The third part is known as the "social" component, which represents the collaboration among the particles. In addition, the implementation of PSO also requires placing a limit on the particle velocity, and the limit, i.e. the maximum allowed velocity V_{max}, determines the searching granularity of space. The inertia weight w plays the role of balancing the global search and local search, and it can be a positive constant or even a positive linear or nonlinear function of time.

V. EXPERIMENTAL SETUP AND VIBRATION DATA

In this section, an experimental dataset of a typical ballbearing is considered. These data are recorded by Tabaszewski and Cempel (1998). The ball bearing type that has been tested is 6402 (steel cage). The shaft rotational frequency and the sampling frequency of the analyzer for recording the acceleration (m/s^2) signal of the ball bearing are 24.5625 and 16,384 Hz, respectively. For the data acquistion, the B&K analyser is used [28].

VI. SYSTEM IMPLEMENTATION DETAILS

The database is composed from five different classes C1 to C5, including a normal bearing and four faults of roller bearing (outer race completely broken fault bearings, broken cage with one loose element fault bearings, damaged cage with four loose elements fault bearings and badly warned ball-bearings). Once the features are extracted by the AR-PCA method, the total database of bearing faults were divided in two sets: one for training (containing 60; 71% of the samples), and the other for test (containing 39; 29% of the samples).

The classification performance of SVM is affected by two techniques, the choosing of the kernel parameters, and the choosing of the regularization parameter c [4].

The proposed approaches for SVM parameter optimization with PSO, is as follows:

Step 1. *Particle initialization and PSO parameters setting:* Set the PSO parameters including: c_1 , c_2 , position of each particle, velocity of each particle, number of particles, number of iterations and velocity limitation .

Step 2. *Fitness evaluation:* Perform SVM on each particle in population and compute the prediction accuracy.

Step 3. Update the global and personal best (P_i and P_G) according to the fitness evaluation results.

Step 4. *Particle manipulations:* Each particle moves to its next position using formula (21) and (22).

Step 5. *Stop condition checking:* If stopping criteria (maximum iterations predefined) are not met, go to **step 2**, otherwise, go to the next step.

Step 6. End the training and testing procedure and save optimal c , δ for SVM.

The swarm size is set to 30 particles. The searching ranges for c and δ are as follows: c ∈ [0; 10], and $\delta \in [0; 10]$. Preliminary experiments also let this study set the personal and social learning factors (c_1, c_2) = (1.3, 1.3) that achieves better classification accuracy.

The inertia weight is set to the following equation:

$$w(k) = w_{max} - \frac{(w_{max} - w_{min})}{k_{max}}.k$$

where w_{max} is the initial weight, w_{min} is the final weight, k_{max} is the maximum number of iterations or generation, and k is the current iteration number. The predefined maximum iteration is 10. When the maximum iteration is reached, the accuracy of test set is calculated by the predicted output of the trained SVM classifier.

VII. ANALYSIS OF EXPERIMENTATION RESULTS

In order to select the optimal values of the parameters p (order of autoregressive modeling), for bearing fault classification, a series of experiments had been carried out by varying the values of this parameter.The important variation range of parameter p is given as follows:

- p from 24 to 48,

The classification results in validation and in test obtained for different values of p are shown in table I, where the best classification result of bearing fault in the validation set 97.06% and in the test set 100% is obtained by using p = 39; ∂ = 0:3093 and c = 7:2086.

Part of eigenvectors and recognition results are presented in table II, where the data with * presents test samples, the others present training data.

Table III gives the classification result for this bearing fault classification problem based on :

- the proposed method (SVM based on autoregressive modeling followed by PCA feature extraction).
- SVM based on wavelet packet feature extraction ,where Discrete wavelet packet transform was used to decompose the time signals into eight packets at level 3 via Daubechies-8.

A confusion matrix of dimension 5 *5 is constructed to show the bearing fault classification performance. The diagonal elements represent the correctly classified bearing fault. The off-diagonal elements represent the misclassification of bearing faults.

We can draw conclusions from experiments:

- when the order of autoregressif modeling (p) increase, It was noted a general trend of increasing of The Validation and test rate.
- the classification accuracy is poor either in validation or test when we combine SVM with wavelet packet(W.P).
- the best classification result of bearing fault in the validation set 97,06% and in the test set 100% is obtained by using the proposed method of SVM-PSO based on Autoregressive Modeling followed by PCA feature extraction, where only 1 damaged cage with four loose elements fault bearing was judged to one loose elements fault bearing by error.

These results clearly show the high percentage of correct classification reached for the validation set and test set, which clearly shows the good generalization capacity of SVM-PSO based on AR modeling and PCA for fault diagnosis of roller bearing.

TABLE I
SVM CLASSIFICATION RESULTS OF BEARING FAULT USING DIFFERENT VALUES OF P.

Ordre of AR Modeling (p)	size of input after applied PCA	Optimal ∂	Optimal c	Rate of validation %	Rate of Test %
24	3×1	0.2969	7.4036	76.47	100
27	3×1	0.1280	4.7472	91.18	95.45
36	3×1	0.1741	8.3113	94.12	95.45
39	4×1	**0.3093**	**7.2086**	**97.06**	**100**
45	4×1	0.3435	8.1661	97.06	100
48	4×1	0.5789	8.1317	91.18	95.45

TABLE II
PART OF EIGENVECTORS AND RECOGNITION RESULTS

input	1	2	3	4	5	6
λ_1	-3.2426	-0.7951	-0.9907	-0.8133	-0.8524	-1.2722
λ_2	-0.6186	-0.1080	0.4806	-0.5672	0.5961	0.5314
λ_3	0.0376	0.7607	0.0011	-0.0164	-0.2214	0.0263
λ_4	-0.3145	0.4669	0.0246	0.0233	0.0410	-0.0497
State	C1	C2	C3	C3*	C4*	C5*
Result	C1	C2	C4	C3*	C4*	C5*

TABLE III
CLASSIFICATION RESULT OF BEARING FAULT IN VALIDATION AND TEST

SVM	Kernel function	Validation results	Test results	Validation Rate(%)	Test rate (%)
AR +PCA	Gaussian	70000 06000 00610 00070 00007	40000 03000 00500 00050 00005	97.06	100
W.P	Polynomial	70000 06000 00511 00250 00106	40000 03000 00140 00014 00014	85.29	63.64

VIII. CONCLUSION

In this paper, an intelligent diagnostic method based on AR-PCA with SVM-PSO approach is presented for fault diagnosis of bearings. The proposed method adopts AR-PCA algorithm to extract features and SVM classifier to identify faults, however a particle swarm optimization is employed to simultaneously optimize the SVM kernel function parameter and the penalty parameter.

The application demonstrates that the proposed approach can greatly improve the accuracy and efficiency of the bearing fault diagnosis.

IX. REFERENCES

[1] M. Li and P. Zhao, “The application of wavelet packet and svm in rolling bearing fault diagnosis,” IEEE International Conference on Mechatronics and Automation, August 2008.

[2] P. K. Kankar, C. S. Satish and S. P. Harsha, “Fault diagnosis of ball bearings using machine learning methods,” ELSEVIER, 2011.

[3] L. Shuang and L. Meng, “Bearing fault diagnosis based on pca and svm,” IEEE International Conference on Mechatronics and Automation, August 2007.

[4] G. Xian, “Mechanical failure classification for spherical roller bearing of hydraulic injection moulding machine using dwt-svm,” ELSEVIER, August 2010.

[5] V. N. Vapnik, The Nature of Statistical Learning Theory. Springer-Verlag, 1995.

[6] ——, Statistical Learning Theory. Springer, 1998.

[7] C. Cortes and V. N. Vapnik, Support-vector networks. Machine Learning, 1995.
[8] C. J. Burges, “A tutorial on support vector machines for pattern recognition,” Data Mining and Knowledge Discovery, vol. 2, pp. 121–167, 1998.
[9] S. R. Gunn, “Support vector machines for classification and regression,” pp. 1–28, 1998.
[10] J. Christopher and C. Burges, “A tutorial on support vector machines for pattern recognition,” Kluwer Academic Publishers, Bell Lab, Lucent Technologies, Boston, pp. 1–43, 1998.
[11] R. Yuan and B. Guangchen, “Determination of optimal svm parameters by using ga/pso,” Journal of computer, vol. 5, no. 8, August 2010.
[12] H. Cheng-Lung and D. Jian-Fan, “A distributed pso-svm hybrid system with feature selection and parameter optimization,” ELSEVIER, 2008.
[13] X. Yun-Jie and X. Shu-Dong, “A new and effective method of bearing fault diagnosis using wavelet packet transform combined with support vector machine,” Journal of computers, vol. 6, no. 11, November 2011.
[14] J. Kennedy and R. C. Eberhart, “Particle swarm optimization,” Proceedings of International Conference on Neural Networks, IEEE, pp. 1942– 1948, 1995.
[15] J Kennedy, R C Eberhart and Y.Shi, “Swarm intelligence,” Morgan Kaufmann Publishers Inc, 2001.
[16] J. LI, “A combination of pso and svm for road icing forecast,” Journal op computers, vol. 5, no. 9, p. 2010, September.
[17] P. Stoica and R. Moses, Spectral analysis of signal. Pearson Education, 2005.
[18] A. Gattal, S. Chenikher, K. M. Pekpe and J. P. Cassar, “Bearing faults diagnosis using artificial network with dimensionality reduction by pca,” 1st International Conference on Electrical Engineering CIGET09, 2009.
[19] L. Wang, Support Vector Machines: Theory and Applications. Springer- Verlag Berlin Heidelberg, 2005.
[20] P. Clarkson and P. J. Moreno, “On the use of support vector machines for phonetic classification,” IEEE Proceedings of the international conference acoustics, speech and signal processes, vol. 2, pp. 585 588, March 1999.
[21] P. Ding, Z. Chen, Y. Liu and B. Xu, “Asymmetrical support vector machines and applications in speech processing,” IEEE Proceedings of the international conference acoustics, speech and signal processes, vol. 1, pp. 73–76, May 2002.
[22] C. Wang and Y. Song, “Support vector machine for mechanical faults diagnosis,” IEEE International Conference on Measuring Technology and Mechatronics Automation, 2010.
[23] B. Scholkopf, A. Smola, R. C. Williamson and P. L. Bartlett, “New support vector algorithms,” Neural Computation, vol. 12, pp. 1207–1245, 2000.
[24] M. XiaoXiao and H. XiYue, “2ptmc classification algorithm based on support vector machines and its application to fault diagnosis,” Control and Decision, vol. 18(3), pp. 272–276, 2003.
[25] J. Weston and C. Watkins, “Multi-class support vector machines,” European Symposium on Artificial Neural Networks, pp. 219–224, April 1999.
[26] X. F. Yuan and Y. N. Wang, “Parameter selection of svm for function approximation based on chaos optimization,” Journal of Systems Engineering and Electronics, vol. 19, pp. 191–197, 2008.
[27] B. Ustun, W. J. Melssen , M. Oudenhuijzen and L. M. C. Buydens, “Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization,” Analytica Chimica Acta, vol. 54, pp. 292–305, 2005.
[28] M. Tabaszewski and C. Cempel, “Ball bearing dataset, structural integrity and damage assessment. network (sidanet.).” available at: www.sidanet.org, 1998. (consulted in July 2006).