Research Article
Hybrid PCA/SVM Method for Recognition of Non-Stationary Time Series

Shao Qiang and Feng Chanjian

Department of Mechanical Engineering, Dalian Nationalities University, Dalian 116600, China

Abstract: A SVM (Support Vector Machine)-like framework provides a novel way to learn linear Principal Component Analysis (PCA), which is easy to be solved and can obtain the unique global solution. SVM is good at classification and PCA features are introduced into SVM. So, a new recognition method based on hybrid PCA and SVM is proposed and used for a series of experiments on non-stationary time series. The results of non-stationary time series recognition and prediction experiments are presented and show that the method proposed is effective.

Keywords: Chatter gestation, pattern recognition, PCA, SVM

INTRODUCTION

In modern and unmanned machining systems, including dedicated transfer lines, flexible manufacturing systems and Reconfigurable Manufacturing Systems (RMS), one crucial component is a reliable and effective monitoring system to monitor process conditions and to take remedial action when failure occurs, or is imminent. Vibration monitoring method is adopted because it is of cheapness and convenience. Hoverer the monitoring vibration signals are usually some non-stationary time series. The Detection and identification on these time series belong to the problem of dynamic pattern (Wang, 2006).

Many techniques in pattern recognition deal with static environments: the class distributions are considered relatively constant as a function of the time in which feature vectors are acquired. However, Time often plays a secondary role: it should be incorporated in the feature extraction procedure. For practical recognition tasks, the assumption of stationarity of the class distributions may not be hold. Alternatively, information in sequences of feature vectors may be used for recognition. We will call them dynamic pattern recognition problems. A dynamic pattern is a multidimensional pattern that evolves as a function of time.

A set of feature vectors can be looked upon as the result of independent draws from a multi-dimensional distribution. All temporal information should present in each feature vector. Identification problem may then be based on the dissimilarity of a set of newly measured feature vectors with respect to a set of known templates.

For running rotor machine, it is necessary to identify the type of faults during it's early stage for the selection of appropriate operation actions to prevent a more severe situation, or to mitigate the consequences of the fault. It is not easy for an operator to identify the type of faults accurately, using the information given by instruments and alarms, with a limited time interval. Therefore, the use of a computer-based Fault diagnosis is recommended. This method is intended to support an operator's decision-making, or to provide input signals fro a computerized faults monitoring system and a computerized operating–procedure management system.

PCA is one of the most widely used tools for learning probabilistic models of dynamical signal series (Lu-Hsien *et al*., 2011). And PCA can model dynamical behaviors variation existing in the system through a latent variable, while SVM shows superior performance in classification. In this study, we seek to cope with above problems by integrate PCA and SVM. Firstly, a pre-processing scheme based on PCA is given to extract good feature from input attributes. Secondly, a SVM scheme for quality classification is provided to classify the chatter data. The experiments results show the proposed method is effective.

THEORETICAL BACKGROUND

Principal component analysis: PCA (Rui and Wenjian, 2011) is a very popular data pre-processing algorithm that provides a lower dimension from a complex dataset and it still effectively retains the characteristics of the data set while having simplified structure and able to reveal underlying features in the dataset. The greatest variance by any projection of the data becomes the first principal component; the second greatest variance becomes the second principal component and so on. The lower order components will be the ones to keep as they retain most of the important

Corresponding Author: Shao Qiang, Department of Mechanical Engineering, Dalian Nationalities University, Dalian 116600, China

aspects of the dataset. Hence, PCA is often used as a preprocessing step to clustering.

The typical processing for PCA are presented in the following (Paulo, 2005; Liao *et al.*, 2007). Consider a set of M stochastic signals $X_i \in \Re^N$, $i = 1,\ldots$ M, each represented as a column vector, with mean $m_x = 1/M \sum_{i=}^{M} X_i$. The purpose of the KL transform is to find an orthogonal basis to decompose a stochastic signal x, from the same original space, to be computed as X = $Uv+m_x$, where the vector $v \in \Re^N$ is the projection of x in the basis, i.e., v = U^T (x-m_x). The matrix U = [u_1 u_2 . . . u_N] should be composed by the N orthogonal column vectors of the basis, verifying the eigenvalue problem:

$$R_{xx}u_j = \lambda_j u_j, \; j = 1,\ldots\ldots, N \tag{1}$$

where, Rxx is the ensemble covariance matrix, computed from the set of M experiments:

$$R_{xx} = \frac{1}{M-1}\sum_{i=1}^{1} \left(X_i - m_x \right)\left(X_i - m_x \right)^T \tag{2}$$

Assuming that the eigenvalues are ordered, i.e., $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$, the choice of the first n << N principal components, leads to an approximation to the stochastic signals given by the ratio on the covariance's associated with the components, i.e:

$$\sum_n \lambda_n / \sum_N \lambda_N$$

In many applications, where stochastic multidimensional signals are the key to overcome the problem at hand, this approximation can constitute a large dimensional reduction and thus a computational complexity reduction. The advantages of PCA are threefold:

- It is an optimal (in terms of mean squared error) linear scheme for compressing a set of high dimensional vectors into a set of lower dimensional vectors
- The model parameters can be computed directly from the data (by diagonalizing the ensemble covariance)
- Given the model parameters, projection into and from the bases are computationally inexpensive operations of complexity O (nN).

SVM for classification: Support Vector Machines, first proposed by Vapnik (1995, 1998), based on Vapnik-Chervonenkis theory and structural risk minimization, is an important tool for machine learning. The main idea of SVM is to first map the data points into a high-dimensional feature space by using a kernel function and then to construct an optimal separating hyperplane between the classes in that space. The primary advantage of SVM over the traditional learning algorithm is that the solution of SVM is always globally optimal and avoids local minima and over-fitting in the training process. For further details on SVMs. The algorithm for gait classification is briefly introduced as follows (Shijie *et al.*, 2012).

Given that a gait data set H of M points in an n-dimensional space containing two different classes +1 and -1 (here +1 represents the elderly and -1 the young):

$$H = \left\{ \left(x_k, y_k \right) \right\}_{k=1}^{M}$$

where,

$$x_k \in R^n, y_k \in \left\{ -1, +1 \right\} \tag{3}$$

The SVM can map a given measurement xi into its label space $y_k \in \{-1, +1\}$:

$$f : R^n \rightarrow \left\{ +1, -1 \right\}, x_i \rightarrow y_i \tag{4}$$

For a test gait data x, the optimal separating hyper plane in SVM is formulated as:

$$f(x) = sign\left(\sum_{i \in SV} \beta_i y_i K(x_i, x) + b \right) \tag{5}$$

where, $K(x_i, x_j)$ is a kernel function satisfying Mercer's conditions, b is a bias estimate in the training process, bi are the coefficients of the generalized optimal separating hyper plane, which are obtained by solving the following quadratic programming problem:

$$\min \; W(\beta) = -\beta^T I + \frac{1}{2} \beta^T H \beta \tag{6}$$

Subject to $\beta^T y = 0$ and $\beta_i \in [0, C`]$
where, $\{\beta\} = \beta_i$, $\{I_i\} = 1$ and $H = y_i y_j K(x_i, x_j)$ i, j = 1,…, *M*.

For the nonlinearly separable gait data, the misclassification penalty parameter C0 can control the trade-off between the maximum margin and the minimum training errors and must be set to a given value in the training process. Similarly, the kernel function is very important for SVM since it defines the nature of the decision surface that classifies gait data. In this study, the following three kernels.

CHATTER RECOGNITION MODEL BASED ON PCA-SVM

This investigation uses the chatter data of cutting from our laboratory. Firstly, the PCA program is used to find the principal component in these features and the programs are developed using MATLAB. And then the
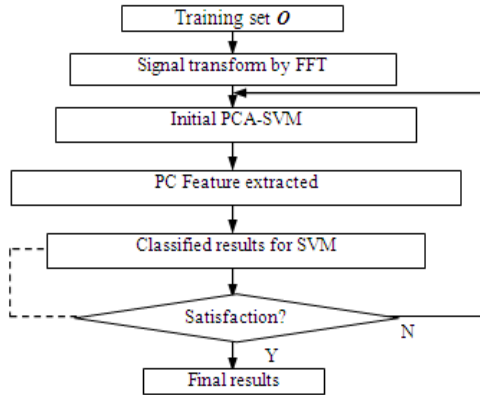
Fig. 1: Hybrid PCA-HMM training

Fig. 2: Experiment equipment

principal components are selected using kernel as the input samples of SVM to solve the site selection problem. By learning and training, we use the data of this subset to find interrelationship of input and output and get the solution by the PCA-SVM mode in Fig. 1.

To develop the PCA detection model:

- Acquire a period of normal building cooling load data and normalize the data using the mean and standard deviation of each variable.
- Choose kernel function $K(x_i, x_j)$, map the original inputs into a high dimensional feature space F.
- Select appropriate number of principal components, develop the PCA model from the scaled data array and calculate the principal component scores.

To develop the SVM forecasting model: Suppose chatter data set for training is $(x_1, y_1), (x_2, y_1), \ldots, (x_n, y_n)$, $x_i \in R$, $y_i \in R$, where n is the number of samples, p is SVM number of input vector (the PCs of chatter data).

- Using the principal components as the input samples of SVM, which regard as xi in training sample?
- Find the optimal solution of Eq. (4) by training data (xi, yi), suited kernel function K(xi, xj) and punishment parameter c. obtain α_i, b and the corresponding support vectors.
- By above conditions and Eq. (4) obtain the cutting chatter forecasting model.

EXPERIMENTS AND RESULTS

Data collection: Experimental data were collected from the drive-end ball bearing of an induction motor driven mechanical system shown in Fig. 2 (Shao *et al.*, 2008). The accelerometer was mounted on the motor housing at the drive end of the motor. Data was collected for four different fault conditions:

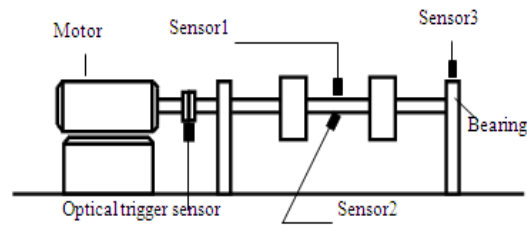- Normal (N)
- Inner Race Fault (IRF)
- Outer Race Fault (ORF)
- Ball Fault (BF)

Faults were introduced into the drive-end bearing by the Electrical Discharge Machining (EDM) method. For the inner race and ball fault cases, vibration data for three severity levels (0.1778, 0.3556 and 0.5334 mm dia) was collected. For the outer race fault case, vibration data for two different severity levels (0.1778 and 0.5334 mm dia) was collected. As the fault diameters suggest, we only considered early damage. The depth of the faults was chosen such that the balls span the gap without bottoming. All the experiments were repeated for four different load conditions (0, 1, 2 and 3 HP). The motor was running directly from the line at approximately 1200 rpm under 0, 1, 2 and 3 HP load.

Figure 3a, b, c and d are the time series extracted from the measured vibration signals from above four experiments of norm, damage at inner race, damage at outer race and damage at ball of bearings separately. In the Fig. 3, x-ordinates substitute the sample points; y-ordinates substitute amplitude of vibrations (Unit mm).

Vibration feature extraction: Linear predictors are used to predict the value of the next sample of a signal as a linear combination of the previous samples. The next sample of the signal $\overline{s}_n$ is predicted as the weighted sum of the p previous samples, $s_{n-1}, s_{n-2}, \ldots, s_{n-p}$, $\overline{s}_n$ can be expressed as:

$$\overline{s}_n = a_1 s_{n-1} + a_2 s_{n-2} + \cdots + a_p s_{n-p} = \sum_{i=1}^{p} a_i s_{n-i} \quad (7)$$

The residual error en is defined as the difference between the actual and predicted values of the next sample and it can be expressed as:

$$e_n = s_n - \overline{s}_n = s_n - \sum_{i=1}^{n} a_i s_{n-i} \quad (8)$$

The weighting coefficients, also referred to as the Linear Prediction Coefficients (LPC) $a_1, a_2, \ldots, a_p$, can be calculated by minimizing some functional of the residual signal en over each analysis window. Different methods can be used to find the linear prediction coefficients. The coefficients of linear predictors are equal to that of AR models.
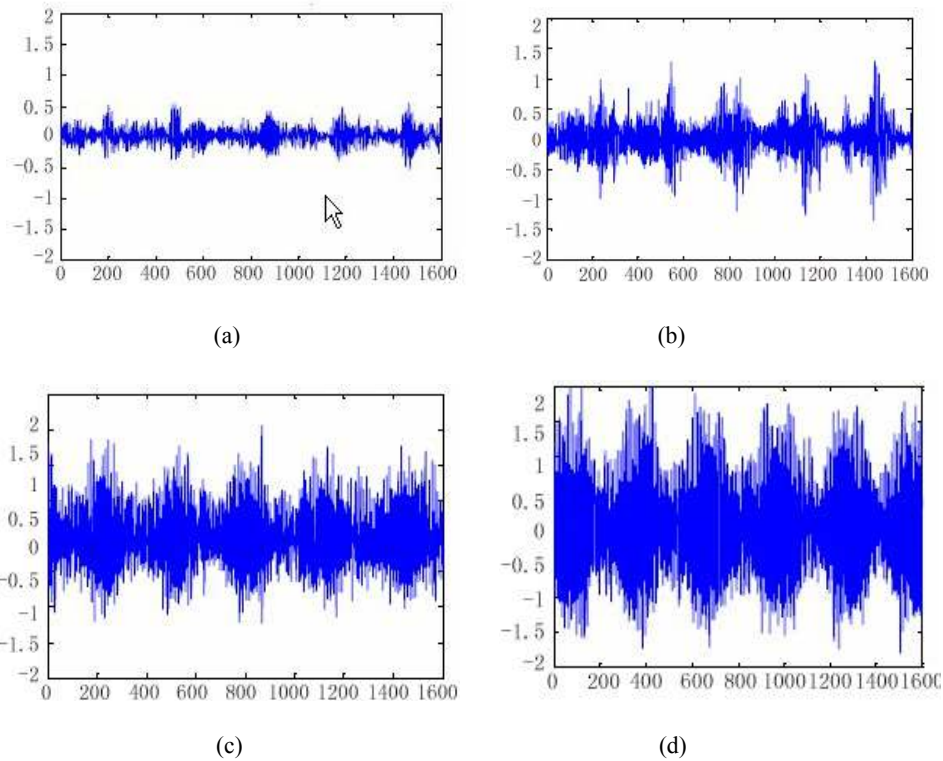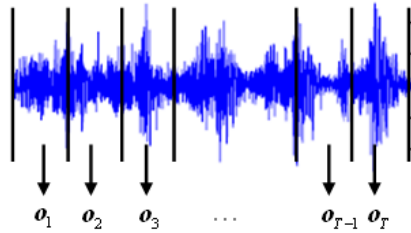
(a)

(b)

(c)

(d)

Fig. 3: Four types time series

Fig. 4: Vibration feature extraction

Table 1: Identification accuracies

LPC orders	Norm (%)	Inner race Image (%)	Outer race Damage (%)	Ball damage (%)
8	94	91	90	93
12	93	93	93	92
16	98	95	93	95
24	100	96	97	100

Vibration signals are non-stationary. Therefore, the future behaviour of a vibration signal is unpredictable. However, when the signal is divided into several small windows, quasi-stationary behaviour can be observed in each window. Thus, future behaviour of the vibration signal can be predicted separately in small windows under the restriction that a different model is used for each window.

In this approach, as illustrated in Fig. 4, the signal is divided into windows of equal length. Each window is coded into a feature vector, which consists of a set of linear prediction coefficients for that window. The feature vectors for all windows are combined together to form a feature matrix. We will interchangeably use observation matrix and feature matrix throughout the rest of the study. In this way, the vibration signal is a feature or observation matrix, which will then be used for training the models.

The observation matrix is $O = [o_1|o_2|o_3 \ldots o_{T-1}|o_T]$, where the o_i is the vector of linear prediction coefficients for i-th window signal.

Identification results: The whole training and test time series was 1600 points samples. The time series is divided into the window of 256 points sample. The feature vector is extracted from the window signal. 8, 12, 16 and 24 orders coefficients are used for features. The identification accuracies are shown in Table 1.

CONCLUSION

In this study, we propose a novel hybrid approach by integrating PCA and SVR for chatter recognition. The original inputs are firstly transformed into nonlinear principal components using PCA. Then these new features are used as the inputs of SVM to solve the cutting chatter recognition. By learning and training, we use the data of this subset to find interrelationship of input and output and get the solution by the PCA-SVM. This method has better convergence ability and strong global search ability, which consumes less time and

better extensive capability than traditional methods on chatter recognition.

REFERENCES

Liao, L.Z., S.W. Luo and M. Tian, 2007. Whitenedfaces" recognition with PCA and ICA. IEEE Signal Proc. Lett., 12: 1008-1011.

Lu-Hsien, C., L. Gwo-Fong and H. Chen-Wang, 2011. Development of design hyetographs for ungauged sites using an approach combining pca: Som and kriging methods. Water Resour. Manage., 2: 1269-1291.

Paulo, O., 2005. PCA positioning sensor characterization for terrain based navigation of UVs. 2nd Iberian Conference on Pattern Recognition and Image Analysis, pp: 615-622.

Rui, Z. and W. Wenjian, 2011. Learning linear and nonlinear PCA with linear programming. Neural Proc. Lett., 33: 151-170.

Shao, Q., S. Cheng and F. Chang-Jian, 2008. Identification of nonstationary time series based on SVM-HMM method. IEEE International Conference on Service Operations and Logistics and Informatics, Inst. of Adv. Control Technol., Dalian Univ. of Technol., Dalian, 1: 293-298.

Shijie, J., G. Yuesheng and Z. Juan, 2012. Product image classification based on local features and svm classifier. JCIT, 7(5): 1-9.

Vapnik, V.N., 1995. The Nature of Statistical Learning Theory. Springer, New York.

Vapnik, V.N., 1998. Statistical Learning Theory. Wiley, New York.

Wang, W., 2006. Research on too1 condition monitoring and on-line compensation technology in milling special spiral rod. Ph.D. Thesis, Northeast University, China. 37.