한국소음진동공학회 2005년도 추계학술대회논문집, pp. 127~131.

Study of Nonlinear Feature Extraction for Faults Diagnosis of Rotating Machinery

Achmad Widodo, Bo-Suk Yang

School of Mechanical Engineering, Pukyong National University, San 100, Yongdang-dong, Nam-gu, Busan 608-739, South Korea

ABSTRACT

There are many methods in feature extraction have been developed. Recently, principal components analysis (PCA) and independent components analysis (ICA) is introduced for doing feature extraction. PCA and ICA linearly transform the original input into new uncorrelated and independent features space respectively. In this paper, the feasibility of using nonlinear feature extraction will be studied. This method will employ the PCA and ICA procedure and adopt the kernel trick to nonlinearly map the data into a feature space. The goal of this study is to seek effectively useful feature for faults classification.

Keywords: Nonlinear Feature Extraction; Kernel; Independent Component Analysis; Principal Components Analysis

1. Introduction

The application of intelligent system for condition monitoring and fault diagnosis is widely used in many areas. In this system, the classification process is needed for condition monitoring and faults diagnosis. For doing good classification process, the preparation of data inputs for classifier needs special treatment to guarantee the good performance in classifier. Many methods have been developed to create the best preparation for data inputs. Recently, the use of feature extraction for data preparation before inputting into classifier has received considerable attention [1]. Feature extraction means transforming the existing features into a lower dimensional space which is useful for feature reduction to avoid the redundancy due to high dimensional data.

Most of feature extraction techniques have based on linear technique such as principal component analysis (PCA) and independent component analysis (ICA). PCA uses a set of basis functions to optimally model the data in sense of minimum error. ICA is relatively recent method that can be considered as generalization of PCA. The ICA method can find a linear transform for the observed data using a set of basis functions where the components are not only decorrelated but also as mutual independent as possible.

The kernel trick is one of the crucial tricks for machine learning. Its basic idea is to project the input data into a high-dimensional implicit feature space with a nonlinear mapping, and then the data is analyzed so that nonlinear relations of the input data can be described. Recently, Bach presented a new learning method of ICA, which use contrast function based on canonical correlation in reproducing a kernel Hilbert space (RKHS) [2]. The other method proposed by Harmeling et al. using a kernel-based blind source separation algorithm in the blind separation of nonlinearly mixed speech signal [3].

In this paper, we introduce the algorithm that incorporates ICA in the kernel trick to improve the feature extraction process that will be used in condition monitoring and faults diagnosis. ICA is formulated in the kernel-inducing feature space and developed through two-phase kernel ICA algorithm: whitened using kernel principal component analysis (KPCA) plus ICA. KPCA spheres data and makes the data structure become as linearly separable as possible by virtue of an implicit nonlinear mapping determined by kernel. ICA seeks the projection direction in the KPCA whitened space, making the distribution of the projected data as non-Gaussian as possible.

2. Kernel Independent Component Analysis (KICA)

2.1 Sphering of data using KPCA

Given an observation sequence $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M$ in R^n, let us assume the features are centered in feature space according to equation

$$\sum_{j=1}^{M} \Phi(\mathbf{x}_j) = 0 \tag{1}$$

The covariance operator, on the feature space can be constructed by

$$\mathbf{S}_t^{\Phi} = \frac{1}{M}\sum_{j=1}^{M} \Phi(\mathbf{x}_j)\Phi(\mathbf{x}_j)^T \tag{2}$$

From Eq. (2), we can get the non-zero eigenvalues that are positive. Let us define matrix **Q** as

$$\mathbf{Q} = [\Phi(\mathbf{x}_1), \ldots, \Phi(\mathbf{x}_M)] \tag{3}$$

Then Eq. (2) can be expressed by

$$S_t^{\Phi} = \frac{1}{M}\mathbf{Q}\mathbf{Q}^T \quad (4)$$

Moreover, we can construct a Gram matrix using Eq. (3) which is their element can be determined by kernel

$$\mathbf{R} = \mathbf{Q}^T\mathbf{Q} \quad (5)$$

$$\mathbf{R}_{ij} = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j) = (\Phi(\mathbf{x}_i)\cdot\Phi(\mathbf{x}_j)) = K(\mathbf{x}_i, \mathbf{y}_j) \quad (6)$$

Denote $\mathbf{V} = (\gamma_1, \gamma_2, ..., \gamma_m)$ and $\mathbf{\Lambda} = diag(\lambda_1, \lambda_2, ..., \lambda_m)$ are eigenvector and eigenvalues of **R** respectively, we can calculate the orthonormal eigenvectors β_j as

$$\beta_j = \frac{1}{\sqrt{\lambda j}}\mathbf{Q}\gamma_j, \quad j = 1,....,m. \quad (7)$$

Then we define matrix **B** as

$$\mathbf{B} = (\beta_1, \beta_2, ..., \beta_m) = \mathbf{QV\Lambda}^{-1/2} \quad (8)$$

The whitening matrix **P** can be derived from Eq. (8) and expressed by

$$\mathbf{P} = \mathbf{B}\left(\frac{1}{M}\mathbf{\Lambda}\right)^{1/2} = \sqrt{M}Q\mathbf{V\Lambda}^{-1} \quad (9)$$

The mapped data in feature space can be whitened by the following transformation

$$\begin{aligned} \mathbf{y} &= \mathbf{P}^T\Phi(\mathbf{x}) \\ &= \sqrt{M}\mathbf{\Lambda}^{-1}\mathbf{V}^T\mathbf{Q}^T\Phi(\mathbf{x}) \\ &= \sqrt{M}\mathbf{\Lambda}^{-1}\mathbf{V}^T[K(\mathbf{x}_1,\mathbf{x}), K(\mathbf{x}_2,\mathbf{x}), ..., K(\mathbf{x}_M,\mathbf{x}) \\ &= \sqrt{M}\mathbf{\Lambda}^{-1}\mathbf{V}^T\mathbf{R}_x \end{aligned} \quad (10)$$

2.2 Processing using ICA

The following task is to find the mixing matrix **W** in the KPCA-transformed space to recover independent components **s** from **y**

$$\mathbf{s} = \mathbf{W}\mathbf{y} \quad (11)$$

There are many algorithms to perform ICA. In this paper, we employ the second order of ICA, proposed by Belouchrani [4] which is adopted in ICALAB toolbox [5].

In summary, the nonlinear feature extraction in this paper performs two phases: whitened process using KPCA and ICA transformation in the KPCA whitened space.

3. Experiment

3.1 Data acquisition

The experiment is conducted using test-rig that consists of motor, pulley, belt, shaft, and fan with changeable blade angle that represents the load, as shown in Fig. 1. Six induction motors of 0.5 kW, 60 Hz, 4-pole were used to create the data. One of the motors is normal condition (healthy), which is considered as a benchmark for comparing with faulty condition. The conditions of faulty motors are described in Fig. 3.

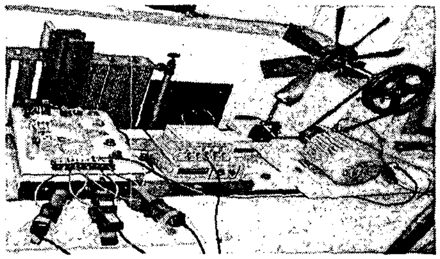

Fig. 1. Test rig and experiment

Three AC current probes and three accelerometers were used to measure the stator current of three phase power supply and vibration signals of horizontal, vertical and axial directions for evaluating the fault diagnosis system. The maximum frequency of the used signals was 5 kHz and the number of sampled data was 16384.

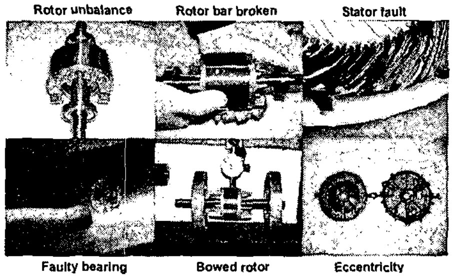

Fig. 2. Faults on the induction motors

3.2 Feature calculation

The total 78 features (13 parameters, 6 signals) are calculated from 10 feature parameters of time domain. These parameters are mean, rms, shape factor, skewness, kurtosis, crest factor, entropy error, entropy estimation, histogram lower and upper. And three parameters from frequency domain (rms frequency, frequency center and root variance frequency) using three direction vibration signals and three–phase current signals. The total of feature parameters can be shown in Table 1.

4. Result and Discussion

Originally, the data feature parameters have disorder structure, fully overlapping and can not be clustered well each condition of faults in induction motors. This phenomenon can be shown in Fig. 3.

Table 1 Data feature parameters

Signals	Position	Feature parameters	
		Time domain	Frequency domain
Vibration	Vertical Horizontal Axial	• Mean • RMS • Shape factor • Skewness • Kurtosis • Crest factor • Entropy error • Entropy estimation • Histogram lower • Histogram upper	• Root mean square frequency • Frequency center • Root variance frequency
Current	Phase A Phase B Phase C		

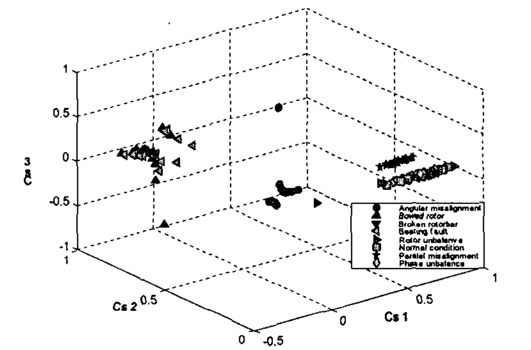

Fig. 3. Three-first components of original data features

Fig. 3 plotted three-first components of original data (78) feature parameters. Because of high dimensional data tends to redundancy and can not be separated well among the condition of faults, so this data structure can not be directly processed into classifier because it will degrade the performance of classifier.

To avoid this disadvantage, we should extract the useful feature and reduce the dimension of original data features. Employing nonlinear feature extraction is expected to be able to handle this condition. In this paper, the use of KPCA for feature extraction also introduced. Based on the eigenvalue, we choose 97% of the total largest eigenvalue of centering kernel matrix as a reference to reduce the dimensionality. Representation of eigenvalue can be seen in Fig. 4 which presents 20 largest eigenvalues of centering kernel matrix. Then we select the RBF kernel function in KPCA and choose the kernel parameter $\sigma = 4$. After feature extraction using KPCA, there are 5 principal components which represent the useful feature. The result of feature extraction using KPCA is presented in Fig. 5.

In Fig.5, we can see that KPCA was successfully clustered each condition of faults in induction motor. However, there are some overlaps in its clustering specially for broken rotor bar and phase unbalance. The good performance of KPCA in clustering is associated that KPCA can explore higher order information of the original data feature beside of uncorrelated data.

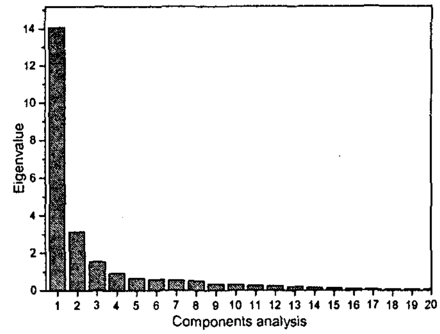

Fig. 4. Representation of 20 eigenvalues of centering kernel matrix

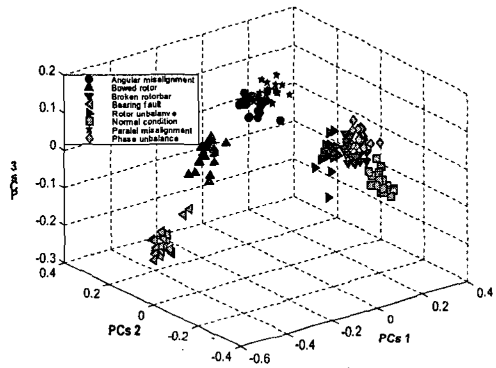

Fig. 5. Feature extraction using KPCA

In the next step, we performed nonlinear feature extraction using whitened data feature by KPCA and employed ICA algorithm to seek the projection direction in KPCA whitened space. We called this process as KICA feature extraction. In this technique, we expect that feature extraction process can be improved due the robustness of KICA. The result of feature extraction using KICA is presented in Fig. 6.

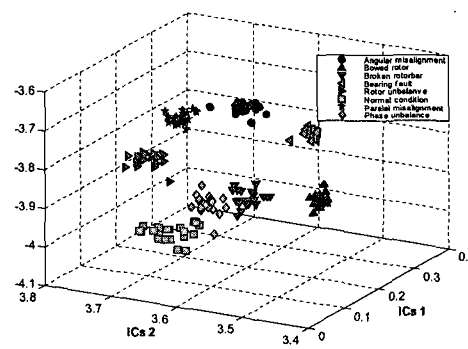

Fig. 6. Feature extraction using KICA

Fig. 6 shows us that each condition of faults in induction motor is separated well. Moreover, there are no overlaps in clustering process. Visually, it can be concluded that feature extraction using KICA is the best in comparing with previous technique. In addition

KICA also implicitly takes into account the high order information of the original data features. Furthermore, in KICA technique, the mutual independent components will give the promising to be a useful and the best features.

To investigate the performance of nonlinear feature extraction process using KPCA and KICA, we calculated the average of Euclidean distance between points in class of feature space [6, 7]. This method can be described as follows: first, we select one point as a reference and calculate the average of Euclidean distance of each point to the reference point. Then we change the reference point and do same as previous step for all data points. We calculated the average of Euclidean distance in KPCA and KICA feature space respectively then took the lowest which represents the good clustering. The calculation of average Euclidean distance can be seen in Fig. 7. In this figure we can see that the average distance of KICA is lower than KPCA so it becomes evidence that performance of KICA significantly outperforms KPCA in terms of clustering.

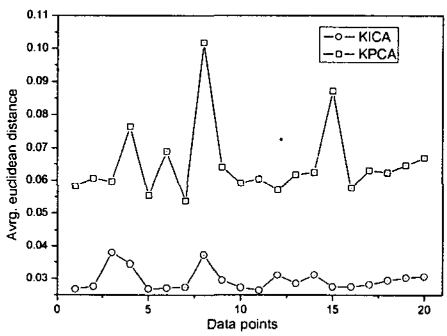

Fig. 7. Average of Euclidean distance KPCA and KICA

5. Conclusion and Future Work

In this paper, ICA formulated in the kernel-inducing feature space and a two-phase kernel algorithm that is KPCA plus ICA is developed. KPCA is used to sphere data feature and to make data as linearly separable as possible using an implicit nonlinear mapping determined by kernel. ICA is followed to seek the projection direction in the KPCA whitened space and determined the mutual components. The effectiveness of KICA in feature extraction is verified using data feature parameters of induction motor. Feature extraction technique using KPCA also introduced to compare with KICA feature extraction process. The result shows that KICA outperforms KPCA in clustering based on the investigation of average of Euclidean distance.

The classification of faults in diagnostic system needs useful feature as a data input in classifier. After get a good features from KICA feature extraction, we will try to train the data features in classifier tool such as support vector machines (SVMs).

Acknowledgement

This work was partially supported by the Brain Korea 21 Project in 2005.

References

[1] Cao LJ, Chua KS, Chong WK, Lee HP & Gu QM. (2003) A comparison of PCA, KPCA and ICA for dimensional reduction in support vector machine. Neurocomputing, 55, 321-336.

[2] Francis R. Bach and Michael I. Jordan. (2002) Kernel independent component analysis. Journal of Machine Learning Research, 3, 1-48.

[3] S. Harmeling, A. Ziehe, M. Kawanabe, B. Blankertz and K. Muller (2001) Nonlinear blind source separation using kernel feature spaces. Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2001), 102-107.

[4] A. Belouchrani, K. Abed-Meraim, J.F. Cardoso, and E. Moulines. (1993) Second-order blind separation of temporally correlated sources, in Proc. Int. Conf. on Digital Sig. Proc., Cyprus, 346-351.

[5] Andrzej Cichocki, Shun-ichi Amari, Krzysztof Siwek, Toshihisa Tanaka, ICALAB for signal processing; Toolbox for ICA, BSS and BSE, http://www.bsp.brain.riken.jp/ICALAB/ICALAB SignalProc/.

[6] M.S. Bartlett, J.R. Movellan, T.J. Sejnowski (2002) Face recognition by independent components analysis. IEEE Trans. Neyral Networks, 13 (6), 1450-1464.

[7] C. Liu, H. Wechsler (2003) Independent component anlalysis of Gabor features for face recognition. IEEE Trans. Neural Networks, 14, (4), 919-928.