한국해양공학회 2006년도 추계학술대회 국제워크샵 pp. 46~53

The Use of Support Vector Machines for Fault Diagnosis of Induction Motors

Achmad Widodo[1,a], Bo-Suk Yang[2,b]

[1]Mechanical Engineering Department, Diponegoro University, Tembalang, Indonesia

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

[a]achmadwid2006@yahoo.com, [b]bsyang@pknu.ac.kr

Keywords: Fault diagnosis; Induction motor; Component analysis; Feature extraction; Support vector machines.

Abstract. This paper presents the fault diagnosis of induction motor based on support vector machine (SVMs). SVMs are well known as intelligent classifier with strong generalization ability. Application SVMs using kernel function is widely used for multi-class classification procedure. In this paper, the algorithm of SVMs will be combined with feature extraction and reduction using component analysis such as independent component analysis, principal component analysis and their kernel (KICA and KPCA). According to the result, component analysis is very useful to extract the useful features and to reduce the dimensionality of features so that the classification procedure in SVM can perform well. Moreover, this method is used to induction motor for faults detection based on vibration and current signals. The results show that this method can well classify and separate each condition of faults in induction motor based on experimental work.

Introduction

The utility of induction motor in industry is unavoidable and becomes important parts to drive their process for producing the products. The issue of robustness and reliability is very important to guarantee the good operational condition. Therefore, faults detection and classification of induction motors has received considerable attention in recent years. Early fault detection and diagnosis can reduce the consequential damage, breakdown maintenance and reduce the spare parts of inventories. Moreover it can increase the prolong machine life, performance, and availability of machine.

Therefore, establishing intelligent system for faults detection of induction motor is very useful and important. To face this issue, the research area in machine learning has been applied to perform condition monitoring, faults detection and classification. One of important issue in faults detection and classification procedure is data preparation. The classifier such as support vector machines (SVMs) will perform well if the data consists of useful features which represent each condition or class. The use of component analysis is aimed to support the data preparation process. Using component analysis, the useful features can be extracted from the original and reduced by removing useless features so that the classifier can reach high accuracy.

Most of feature extraction techniques have based on linear technique such as principal component analysis (PCA) and independent component analysis (ICA). In this paper, ICA and PCA are employed to extract and reduce the features by linear mapping from input data space into feature space. 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 [1]. Furthermore, feature extraction using nonlinear technique [2], namely kernel ICA (KICA) and kernel PCA (KPCA) also introduced in this paper. These are aimed to search better performance in feature extraction than the liner technique one. Nonlinear technique is employed by introducing kernel trick which is crucial trick in 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.

Here, we employed 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 (kernel PCA) plus ICA. Kernel PCA 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 kernel PCA whitened space, making the distribution of the projected data as non-Gaussian as possible.

Finally, after the good features are obtained from feature extraction process then SVMs based on multi-class classification is employed to show the performance of classification.

Preliminary Knowledge

Component analysis. In this section, some basic knowledge of component analysis and SVMs used in this paper will be introduced.

Principal component analysis (PCA). PCA is well known for capturing the linear relationship among feature vectors. Given a set of centered input vectors $\mathbf{x}_t$ ($t = 1, \ldots, l$ and $\sum_{t=1}^{l} \mathbf{x}_t = 0$), each of which is of m dimension $\mathbf{x}_t = (x_t(1), x_t(2), \ldots, x_t(m))^T$ usually $m < l$, PCA linearly transforms each vector $\mathbf{x}_t$ into a new one $\mathbf{s}_t$ by

$$\mathbf{s}_t = \mathbf{U}^T \mathbf{x}_t \tag{1}$$

where $\mathbf{U}$ is the $m \times m$ orthogonal matrix whose ith column, $\mathbf{u}_i$ is the eigenvector of the sample covariance matrix

$$\mathbf{C} = \frac{1}{l} \sum_{t=1}^{l} \mathbf{x}_t \mathbf{x}_t^T \tag{2}$$

In other words, PCA firstly solves the eigenvalue problem

$$\lambda_i \mathbf{u}_i = \mathbf{C}\mathbf{u}_i, \; i = 1, \ldots, m \tag{3}$$

where λ_i is one of the eigenvalues of $\mathbf{C}$, $\mathbf{u}_i$ is the corresponding eigenvector. Based on the estimated $\mathbf{u}_i$, the components of $\mathbf{s}_t$ are then calculated as the orthogonal transformations of $\mathbf{x}_t$

$$\mathbf{s}_t(i) = \mathbf{u}_i^T \mathbf{x}_t, \; i = 1, \ldots, m \tag{4}$$

The new components are called principal components (PCs). By using only the first several eigenvectors sorted in descending order of the eigenvalues, the number of PCs in $\mathbf{s}_t$ can be reduced. So PCA has the dimensional reduction characteristic [3].

Independent component analysis (ICA). ICA is a technique that transform multivariate random signal into a signal having components that are mutually independent in complete statistical sense [4]. Recently this technique has been demonstrated to be able to extract independent components (ICs) from the mixed signals. A generic ICA model can be written as

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{5}$$

where **A** is an unknown full-rank matrix, called the mixing matrix, and **s** is the IC data matrix, and **x** is the measured variable data matrix. The basic problem of ICA is to estimate the IC matrix **s** or to estimate the mixing matrix **A** from the measured data matrix **x** without any knowledge of **s** or **A**. The practical problem of ICA is to calculate a separating matrix **W** so that the components of the reconstructed data matrix $\hat{\mathbf{s}}$ become as independent of each other as possible, given as

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x} \tag{6}$$

In the present work, a fast fixed-point algorithm [5] is used to minimize or maximize the fourth order cumulant to perform ICA.

Kernel PCA (KPCA). KPCA is one approach of generalizing linear PCA into nonlinear case using the kernel method [6]. The idea of KPCA is to firstly map the original input vectors $\mathbf{x}_t$ into a high-dimensional feature space $\varphi(\mathbf{x}_t)$ via nonlinear function called kernel function and then calculate the linear PCA in $\varphi(\mathbf{x}_t)$. The full description of KPCA can be found in reference above.

Kernel ICA (KICA). Practically speaking, the KICA is the combination of centering and whitening process using KPCA as previously explanation and iterative section using ICA. The following task is to find the mixing matrix **W** in the KPCA-transformed space to recover ICs **s** from **r**, recall Eq. (6)

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{r} \tag{7}$$

There are many algorithms to perform ICA. In this paper, we employ a fast fixed-point algorithm [5] which is applied in FastICA toolbox. In summary, the KICA in this paper performs two phases: whitened process using KPCA and ICA transformation in the KPCA whitened space.

Support vector machines (SVMs). SVMs are a kind of machine learning based on statistical learning theory. The basic idea of applying SVM to pattern classification can be stated as follows: first, map the inputs vectors into one features space, possible in higher space, either linearly or nonlinearly, which is relevant with the kernel function. Then, within the feature space from the first step, seek an optimized linear division, that is, construct a hyperplane which separates two classes. It can be extended to multi-class. SVMs training always seek a global optimized solution and avoid over-fitting, so it has ability to deal with a large number of feature. A complete description about SVMs is available in [7]. In the linear separable case, there exists a separating hyperplane whose function is

$$\mathbf{w}\cdot\mathbf{x} + b = 0 \tag{8}$$

which implies

$$y_i(\mathbf{w}\cdot\mathbf{x} + b = 0) \geq 1, \qquad i = 1,\dots, N \tag{9}$$

By minimizing $||\mathbf{w}||$ subject to this constrain, the SVMs approach tries to find a unique separating hyperplane. Here $||\mathbf{w}||$ is the Euclidean norm of **w**, and the distance between the hyperplane and the nearest data points of each class is $2/||\mathbf{w}||$. By introducing Lagrange multipliers α_i, the SVMs training procedure amounts to solving a convex quadratic problem (QP). The solution is a unique globally optimized result, which has the following properties

$$w = \sum_{i}^{N} \alpha_i y_i \mathbf{x}_i \tag{10}$$

Only if corresponding $\alpha_i > 0$, these $\mathbf{x}_i$ are called support vectors.

When SVMs are trained, the decision function can be written as

$$f(\mathbf{x}) = sign\left(\sum_{i=1}^{N} \alpha_i y_i (\mathbf{x} \cdot \mathbf{x}_i) + b\right) \tag{11}$$

For a linear non-separable case, SVMs perform a nonlinear mapping of the input vector $\mathbf{x}$ from the input space $\Re^d$ into a higher dimensional Hilbert space, where the mapping is determined by kernel function. According to the different classification problems, the different kernel function can be selected to obtain the optimal classification results.

Application in Induction Motor

Data acquisition. Data acquisition 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. The motors are conditioned to simulate the faults such as broken rotor bars, bowed rotor, faulty bearing, rotor unbalance, eccentricity and phase unbalance. One of the motors is normal condition, which is considered as a benchmark for comparing with faulty condition. Three AC current probes and three accelerometers were used to measure the steady state, transient 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.

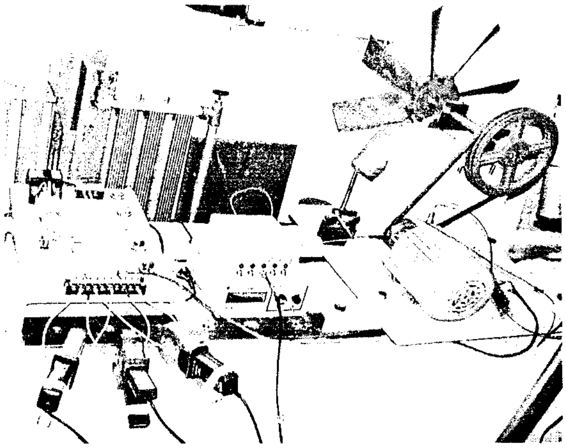


Fig. 1. Test rig and experiment

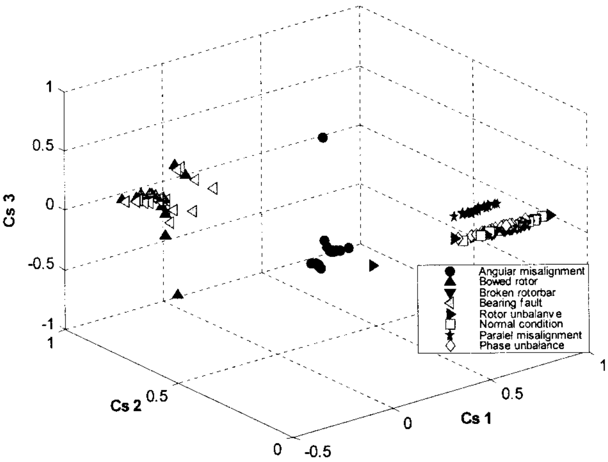


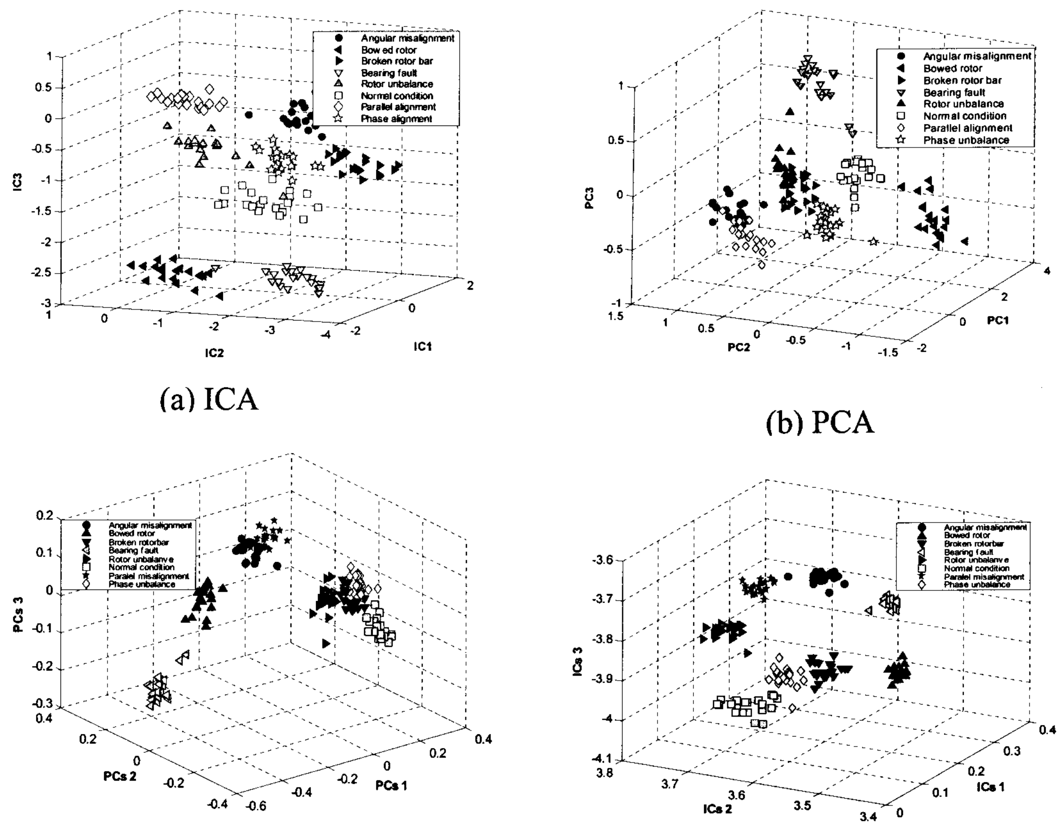
Fig. 2. Original features

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.

Feature extraction and reduction. Originally, the data feature parameters have disorder structure, overlap and each condition of faults in induction motors can not be clustered well. Fig. 2 plots 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 linear feature extraction using ICA and PCA is intended to avoid the disorder structure of data features. In this study, ICA and PCA is applied based on 95% variation of eigenvalue. The first three independent and principal components are plotted in Fig. 3.

It can be observed that the cluster for eight conditions is well separated. Nevertheless, the performance of ICA seems better than PCA does in clustering of each condition.



(a) ICA

(b) PCA

(c) KPCA

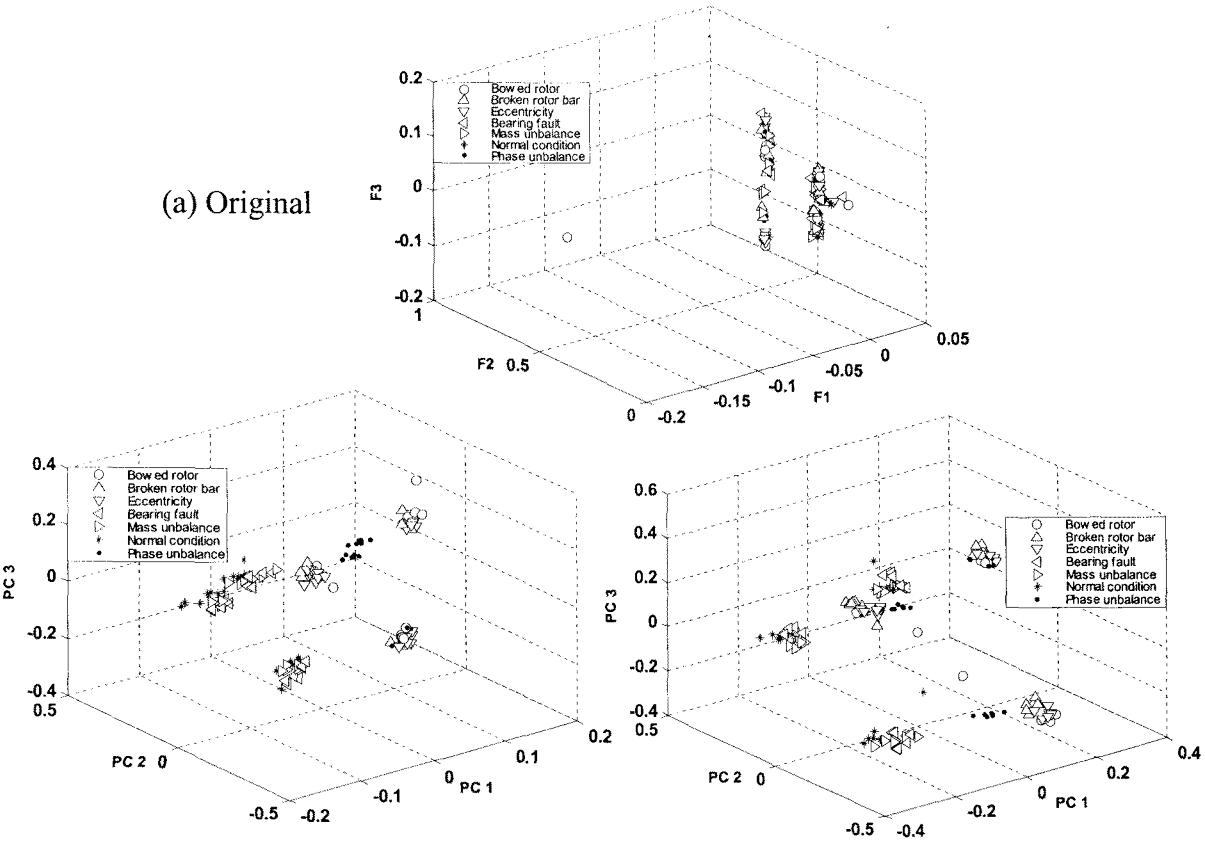
(d) KICA

Fig. 3. Feature extraction

In addition, employing nonlinear feature extraction is expected to be able to handle the disorder structure of data features. In this paper, the use of kernel PCA 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. Then we select the RBF kernel function in kernel PCA and choose the kernel parameter σ = 4. After feature extraction using kernel PCA, there are 7 principal components which represent the useful feature. The result of feature extraction using kernel PCA and kernel ICA is also presented in Fig. 3.

In addition, the faults diagnosis of induction motor is also performed using motor current stator analysis (MCSA) based on transient current signal. Many methods have been developed in MCSA to perform condition monitoring and faults detection of induction motors. A clear brief review discuss about how to use MCSA was highlighted in [8-9].

Before calculating the features, the transient signal need to be preprocessed using smoothing and discrete wavelet transform (DWT) to highlight the salient differences between conditions of induction motor. Furthermore, feature extraction and reduction using PCA and KPCA are employed to avoid the disorder structure of original features such as depicted in Fig. 4a. The results of feature extraction can be seen in Figs. 4b and 4c, however, it can be observed that the clusters for seven conditions are not well separated. There are still overlapping among each condition of motor. It becomes indication that the features which produced by current signature is very difficult to cluster and it needs more advance and good preprocessing so that the salient differences features can be explored and emerged.



(b) PCA (c) KPCA

Fig. 4. Feature extraction of transient current signal

Training and classification. The SVM based multi-class classification is applied to perform the classification process using one-against-one and one-against-all methods. The tutorial of these methods has clearly explained in [10]. To solve the SVM problem, Vapnik [11] describe a method which used the projected conjugate gradient algorithm to solve the SVM-QP problem. Sequential minimal optimization (SMO) proposed by Platt [12] is a simple algorithm that can be used to solve the SVM-QP problem. This method decomposes the overall QP problem into QP sub-problem using the Osuna's theorem to ensure the convergence.

In this study, RBF kernel ($K= \exp(- \|\mathbf{x} - \mathbf{x}_j\|^2 /2\gamma^2)$ and polynomial ($K = (\gamma\, \mathbf{x}^T\cdot\mathbf{x}_j + r)^d$) are used as the basic kernel function of SVMs. There are two parameters associated with these kernels: C and γ. In addition, polynomial kernel also has parameter d related to degree of polynomial. The upper bound C for penalty term and kernel parameter γ play a crucial role in performance of SVMs. Therefore, improper selection of parameters C, γ, and d can cause overfitting or underfitting problem. Nevertheless, there is simple guideline to choose the proper kernel parameters using cross-validation that suggested by Hsu [10].

Beside of two previous kernel function, recently we proposed wavelet-SVM (W-SVM) to perform classification process of transient signal [13]. The aim of using W-SVM is to improve and to obtain better performance in classification routine than the traditional SVM. In this method, wavelet function is performed as kernel function and then induced in SVM theory [14-15].

Results and Discussions

The result of this study can be shown in Tables 1, 2, 3 and 4. In these tables we listed the kernel function, strategy of multi-class classification, classification rate for training and testing, number of support vector and training time. The classification rate (%) is determined by using ratio of correct classification and on the whole of training or testing respectively.

Table 1 Fault classification using PCA and SVMs

Kernel	Multi-class strategy	Classification rate (%) Training	Testing	Number of SVs	Training time (s)
Polynomial	One-against-all (1, 1, 1)	100	100	91	0.063
(d, C, γ)	One-against-one (1, 2^2, 1)	100	100	47	0.031
RBF	One-against-all (1, 2^{-2})	100	100	80	0.063
(C, γ)	One-against-one (1, 2^{-1})	100	99.97	71	0.016

Table 2 Fault classification ICA and SVMs

Kernel	Multi-class strategy	Classification rate (%) Training	Testing	Number of SVs	Training Time (s)
Polynomial	One-against-all (1, 2, 1)	100	100	79	0.062
(d, C, γ)	One-against-one (1, 2^5, 1)	100	100	42	0.031
RBF	One-against-all (2, 1)	100	100	79	0.063
(C, γ)	One-against-one (2, 1)	100	100	64	0.015

Table 3 Fault classification using KPCA and SVMs

Kernel	Multi-class strategy	Classification rate (%) Training	Testing	Number of SVs	Training Time (s)
Polynomial	One-against-all (3, 2^4, 1)	100	100	67	1.33
(d, C, γ)	One-against-one (3, 2^2, 1)	100	100	68	0.031
RBF	One-against-all (2^7, 2^{-2})	100	100	52	0.438
(C, γ)	One-against-one (2^7, 2^{-1})	100	100	50	0.032

Table 4 Fault classification using KICA and SVMs

Kernel	Multi-class strategy	Classification rate (%) Training	Testing	Number of SVs	Training Time (s)
Polynomial	One-against-all (1, 2^6, 1)	100	99.97	67	1.156
(d, C, γ)	One-against-one (1, 2^7, 1)	100	100	50	0.047
RBF	One-against-all (2^6, 2^{-1})	100	100	43	0.218
(C, γ)	One-against-one (2^7, 2^{-2})	100	100	42	0.031

As general, the classification results reached high accuracy both training and testing. By using ICA and PCA feature extraction, the useful feature is extracted from original feature sets. In this case, classification process using ICA feature extraction needs less numbers of SVs than PCA feature extraction.

Nonlinear feature extraction using KPCA and KICA also produced high performance in classification. The best method is obtained when KICA employed in feature extraction process. In comparing with feature extraction using kernel PCA the performance of this method is better according to the number of SVs and training time. This phenomenon exists because KICA performs two phases: whitened process using kernel PCA and ICA transformation in the KPCA whitened space. So that KICA can find the components which is not merely uncorrelated but independent. Independent components are more useful for classification rather than uncorrelated components. Moreover, the other reason is the negentropy in ICA could take into account the higher order information of the original inputs better than PCA using sample covariance matrix.

In the case of transient signal, classification result using W-SVM is presented in Table 5. The accuracy of W-SVM is better than conventional SVM for all wavelet function. W-SVM can recognize well each condition of induction motor even though transient signal has similarity which is difficult to classify using conventional SVM.

Table 5 Classification using W-SVM

		W–SVM			Conventional SVM	
		Haar	Daubechies	Symlet	Gaussian (γ = 0.25)	Poly (d = 3)
Accuracy (%)	PCA	85/85	100/100	100/100	75/75	61/61
(training/test)	KPCA	95/95	100/100	100/100	90/90	74/74
Number of SV	PCA	68	68	68	70	70
	KPCA	68	68	68	70	70
CPU time (s)	PCA	5.8	7.1	9.6	0.9	0.9
	KPCA	5.8	10.9	15.8	0.9	0.8

Conclusions

This paper studies a method for fault diagnosis of induction motor using SVMs and component analysis. The vibration, steady-state and transient current signal are selected as a features source for fault diagnosis. Component analysis using linear method (ICA and PCA) and nonlinear method (KPCA and KICA) are employed for feature extraction and reduction. In addition, a relatively new method called W-SVM is proposed to improve the faults diagnosis based on transient current signal. The results showed that this method is performed well and reached high accuracy in training and testing process based on experimental work.

References

[1] Widodo A., Yang B.S., Han T., Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors, *Expert Systems with Applications* (in press).

[2] Widodo A., Yang B.S., Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors, *Expert Systems with Applications* (in press).

[3] Jolliffe I.J., *Principal Component Analysis*, New York, Springer, (1986).

[4] Back, A.D., Weigend, A.S., (1998). A first application of independent components analysis to extracting structure from stock returns, *International Journal of Neural System 8* (4), 473-484.

[5] Hyvärinen A., Oja E., Independent component analysis: algorithms and applications, *Neural Networks* 13 (4-5), (2000) 411-430.

[6] Scholkopf, B., Smola, A., Muller, K.R., Nonlinear component analysis as a kernel eigenvalue problem, *Neural Computation* 10 (5), (1998) 1299-1319.

[7] Vapnik, V.N., *The Nature of Statistical Learning Theory*, New York, Springer, (1999).

[8] Benbouzid, M.E.H., A review of induction motors signature analysis as a medium for faults diagnosis, IEEE Transaction on Industrial Applications 47 (5) (2000) 984-993.

[9] Douglas, H., Pillay, P., The impact of wavelet selection on transient motor current signature analysis, in Proceedings of IEEE International Conference on Electrical Machine and Drives (2005) 80-85.

[10] Hsu C.W., Lin C.J., A comparison of methods for multiclass support vector machines, *IEEE Transaction on Neural Network*, 13(2) (2002), pp. 415-425.

[11] Vapnik V.N., *Estimation Dependences Based on Empirical Data*, Berlin, Springer Verlag, (1982).

[12] Platt J., Fast training of support vector machines using sequential minimal optimization, in: B. Scholkopf, et al. *Advances in Kernel Methods-Support Vector Learning*, Cambridge, MIT Press, (1999).

[13] Widodo A., Yang B.S., Son J.D., Intelligent fault diagnosis system of induction motor based on transient current signal, *Pattern Recognition Letters* (submitted)

[14] Rakotomamonjy A., Canu S., Frame, reproducing kernel, regularization and learning, *Journal of Machine Learning Research* 6(9), (2005)1485-1515.

[15] Gao, J., Chen F., Shi, D., On the construction of support wavelet network, *IEEE International Conference on System, Man, and Cybernetics*, (2004) 3204-3207.