Thesis for the Degree of Doctor of Philosophy

Support Vector Machine for Machine Fault Diagnosis and Prognosis

by

Achmad Widodo

Department of Mechanical Engineering,

The Graduate School

Pukyong National University

August 2007

Support Vector Machine
for Machine Fault Diagnosis and Prognosis
기계 결함 진단과 예지를 위한 SVM

Advisor: Prof. Bo-Suk Yang

\

by

Achmad Widodo

A thesis submitted in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in the Department of Mechanical Engineering,
The Graduate School
Pukyong National University
August 2007

Support Vector Machine
for Machine Fault Diagnosis and Prognosis

A dissertation

by

Achmad Widodo

Approved by:

Chairman: Prof. Dong-Jo Kim

Member: Prof. Soo-Jong Lee

Member: Prof. In-Pil Kang

Member: Dr. Dong-Soo Lim

Member: Prof. Bo-Suk Yang

June 8, 2007

Contents

List of Figures

List of Tables

List of Symbols

Chapter II Preliminary Review and Study

a	transform scale
a_i	auto-regression coefficient
a_{jk}	approximate coefficient of wavelet
b	translation parameters
c_1	mean
c_2	standard deviation
c_3	skewness
c_4	kurtosis
CF	crest factor
CWT	continuous wavelet transform
d_{jk}	detail coefficient of wavelet
DWT	discrete wavelet transform
$E\{\cdot\}$	expected value
$E_s(x)$	entropy estimation
$E_e(x)$	standard error
FC	frequency center
$g[k]$	low-pass filter
$h[k]$	high-pass filter
h_L	lower bound of histogram
h_U	upper bound of histogram
H_k	coefficient of scaling function
j	scaling coefficient
k	translation coefficient
m_n	moment coefficient

MSF	mean square frequency
N	number of data
$RMSF$	root-mean square frequency
RVF	root-variance frequency
SF	shape factor
VF	variance frequency
x_i	time historical data
$y[n]$	decomposition coefficient

Greek symbols

$\phi(t)$	scale function
$\psi_{a,b}$	mother wavelet
$\psi(t)$	wavelet function

Chapter III Component Analysis and Support Vector Machine

$\mathbf{A}$	mixing matrix
b	bias
$\mathbf{B}$	orthogonal matrix
C	penalty constant
D	dilatation operator
E	expectation
f	function
$F[\omega]$	Fourier transform
G	non-quadratic function
H	entropy
J	negentropy
k	order number

K	kernel function
L	Lagrangian function
m	number of principal component
M	number of input
N	number of data sample
p	dimensional data
$\mathbf{P}$	whitening matrix
$\mathbf{R}$	Gram matrix
$\mathbf{s}$	independent component
T	translation operator
$\mathbf{T}$	linear transformation matrix
$\mathbf{V}$	eigenvector matrix
$\mathbf{w}$	weight vector
$\mathbf{W}$	separating matrix
$\mathbf{x}$	input vector
y_i	class label

Greek symbols

α	Lagrange multiplier
$\boldsymbol{\beta}$	orthonormal eigenvector
ϕ	feature space
γ	kernel parameter
$\boldsymbol{\Lambda}$	eigenvalue matrix
λ_i	eigenvalue
μ	mean data sample
$\hat{\mu}$	global mean of sample
$\boldsymbol{\Sigma}$	diagonal matrix
υ	Gaussian variable zero mean

ξ	slack variable
$\mathbf{\Psi}$	singular vector matrix

Chapter IV SVM Based Fault Diagnosis for Induction Motors

B	ball roller diameter
$BPFO$	ball pass frequency of outer race
$BPFI$	ball pass frequency of inner race
BSF	ball spin frequency
C	penalty term
d	degree of polynomial
D	average distance
f	frequency
f_b	broken rotor bars frequency
f_{brg}	bearing frequency
f_e	supply frequency
f_{ec}	eccentricity fault frequency
f_i	character frequency of inner race
f_o	character frequency of outer race
f_r	rotating frequency
f_t	turn-fault frequency
f_v	characteristic vibration frequency
FTF	fundamental train frequency
n	rotor speed
n_s	synchronous speed
P	pitch diameter
p	number of pole
q	joint feature sets

s	slip

Greek symbols

α	condition patterns
δ	evaluation criterion
γ	kernel parameter
λ	parameter for quadratic-programming
ρ	mean all features

Chapter V Feasibility of SVM for Machine Prognosis System

b	bias
C	penalty constant
Cov	covariance
f	function
K	kernel function
N	number of data
R	correlation coefficient
$RMSE$	root-mean square error
$\mathbf{w}$	weight vector
$\mathbf{x}$	input vector
$\mathbf{y}$	target value
y	observed value
$\hat{y}$	predicted value

Greek symbols

α	Lagrange multiplier
γ	kernel parameter

ϕ	feature space
ξ	slack variable
σ	standard deviation
τ	time delay

Support Vector Machine for Machine Fault Diagnosis and Prognosis

Achmad Widodo

Department of Mechanical Engineering, The Graduate School,
Pukyong National University

Abstract

Recently, the issue of machine fault diagnosis and prognosis as a part of maintenance system became global due to the potential advantages to be gained from reduced maintenance costs, improved productivity and increased machine availability. Numerous methods have been developed based on intelligent system such as artificial neural network, fuzzy expert system, condition-based reasoning, random forest, etc. However, the use of support vector machine (SVM) for machine fault diagnosis and prognosis is still rare. SVM has an excellence performance in generalization so it can produce high accuracy in classification and prediction for machine fault diagnosis and prognosis, respectively.

In this paper, SVM will be redeveloped to be an intelligent system for conducting fault diagnosis and prognosis of machine. SVM has two excellent abilities in the framework of machine learning, those are classification and regression. Fault diagnosis is performed using classification ability of SVM, while the prognosis of machine condition is conducted based on regression using SVM. As an intelligent technique, SVM can train the given data and save the result as weights, and then use the weights for doing classification and regression. Originally, SVM is used for two class classification of linear data; however, using

kernel mapping SVM can perform training process and doing classification with nonlinear data. By optimizing the hyperplane, SVM tries to solve the classification and regression problems.

In the developed system, SVM is combined by technique so-called feature-based technique to do classification for fault diagnosis purpose. Feature-based technique is an effort to represent the raw data as feature such as characteristic values (statistical), color, shape and so on. In machine fault diagnosis, features are representative of values which indicate the machine condition. Using feature, the problem with data transferring and data storage can also be solved. Feature-based classification technique consists of data acquisition, preprocessing, feature representation, feature calculation, feature selection and classifiers. In this study, SVM is adopted, redeveloped and combined with feature-based technique to obtain a novel fault diagnosis tool.

The proposed method is validated using induction motor data to perform fault diagnosis by means of classification strategy in SVM. Several case studies have been done to diagnose fault occurrence in induction motor such as bent rotor, broken rotor bars, bearing fault, mass unbalance, phase unbalance and eccentricity fault. The data used in the experiments are vibration and current data. The results show that the proposed method can perform fault diagnosis well, and it can be concluded that the proposed method may serve the fault diagnosis technique in the future.

Prognosis can be defined as the ability to predict accurately and precisely the remaining useful lifetime of a failing machine component or subsystem. Therefore, a reliable predictor is very important and it is very useful to a wide range of industries to forecast the upcoming states of a dynamic system or to predict damage propagation trend in rotating machineries. In mechanical system, for example, the forecasting information can be used for condition monitoring to provide an accurate alarm before fault reaches critical levels so as to prevent machinery performance degradation, malfunction or catastrophic failure. Moreover, it can be used for scheduling of repairs and predictive/preventive maintenance in manufacturing facilities; and predictive and fault-tolerant control.

In this study, SVM based regression is redeveloped to be a predictor of time-series data. Trending data of machine can be considered as time-series, it contains information of machine during its operation. The proposed method is addressed to predict the upcoming state of machine based on previous condition. Trending data of a low methane compressor is used to validate the proposed method. Performance of prediction is measured using *RMSE* and coefficient correlation (*R*). The result show that SVM based regression has potential and promising to be a reliable prognosis tool.

I. Introduction

1. Background

Since the maintenance has significant impact in industry, it has received a deep attention from the expert and practical maintenance. According to study, maintenance costs are a major part of the total operating costs of all manufacturing and production plants, which can make or break a business. Depending on the specific industry, maintenance costs can represent from 15% to 40% of the costs of goods produced [1]. In fact, these costs are associated with maintenance labor and materials and are likely to go even higher in the future with the addition of factory automation through the development of new technologies.

Nowadays, the development of maintenance strategy was supported by computer technology both in hardware and software. A recent developed method is using artificial intelligent (AI) techniques as tool for maintenance routine. Based on the idea how to perform an excellent and easy maintenance program; it leads the practical maintenance to create an intelligent maintenance system. Intelligent maintenance must consist of parts (hardware and software) which are possible the system to do maintenance routine in such a way like human being. Application of expert system (ES) as a branch of AI in maintenance is one of solutions. The basic idea of ES is simply that expertise, which is the vast body of task-specific knowledge, is transferred from a human to a computer. This knowledge is then stored in the computer and users call upon the computer for specific advice as needed. The computer can make inferences and arrive at a specific conclusion. Then, like human consultant, it gives advice and explains, if necessary, the logic behind the advice [2].

In recent year, it can be said that approximately half of all operating costs in most processing and manufacturing operations can be attributed to maintenance. This is ample motivation for studying any activity that can potentially lower these costs. Machine condition monitoring and fault diagnosis is one of these activities.

According to Williams et al. [3], adopted from British Standard (BS 3811:1984), condition monitoring is defined as the continuous or periodic measurement and interpretation of data to indicate the condition of an item to determine the need for maintenance. Condition monitoring is needed for guarantee the survival of machine so that incipient fault can be detected and diagnosed as early as possible. The possibility of failure cannot be avoided in the machine, but early diagnosis of incipient failure is useful to avoid the machine breakdown. When fault occurrence exists in the machines, it will give some symptoms like excessive vibration and noise, extremely increased temperature, oil debris, etc. Using machine condition monitoring, these symptoms can be early detected and efforts to overcome the breakdown of machine can be realized soon.

Machine condition monitoring and fault diagnosis can also be defined as the field of technical activity in which selected physical parameters, associated with machinery operation, are observed for the purpose of determining machinery integrity [4]. Once the integrity of a machine has been estimated, this information can be used for many different purposes. Loading and maintenance activities are the two main tasks that link directly to the information provided. The ultimate goal in regard to maintenance activities is to schedule only what is needed at a time, which results in optimum use of resources. Having said this, it should also be noted that condition monitoring and fault diagnosis practices are also applied to improve end product quality control and as such can also be considered as process monitoring tools.

There are several benefits and advantages in machine condition monitoring and fault diagnosis as follows

1. Increased machine availability and reliability.
2. Improved operating efficiency.
3. Improved risk management (less downtime).
4. Reduced maintenance costs (better planning).
5. Reduced spare parts inventories.
6. Improved safety.
7. Improved knowledge of the machine condition (safe short-term overloading of machine possible).
8. Extended operational life of the machine.
9. Improved customer relations (less planned/unplanned downtime).
10. Elimination of chronic failures (root cause analysis and redesign).
11. Reduction of post overhaul failures due to improperly performed maintenance or reassembly, etc.

By considering the importance and benefits of machine condition monitoring and fault diagnosis, this research proposes an intelligent machine fault diagnosis system based on support vector machine (SVM). SVM is a relatively new computational learning method based on the statistical learning theory; can serve as ES to carry out intelligent machine condition monitoring system. Introduced by Vapnik and his co-workers [5-7], SVM becomes famous and popular in machine learning community due to the excellence of generalization ability than the traditional method such as neural network. Therefore, SVM have been successfully applied to a number of applications ranging from face detection, verification, and recognition, object detection and recognition, handwritten character and digit recognition, text detection and categorization, speech and speaker verification, recognition, information and image retrieval, prediction and so on. However, research and published papers which discuss the use of SVM in machine condition monitoring, fault diagnosis and prognosis are much fewer.

Therefore, this research is aimed to give a contribution for developing an intelligent method in machine fault diagnosis and prognosis based on SVM.

2. Motivation and Significance of This Research

Machines are critical part in industry. Industrial machines are complex and consist of many components that could potentially fail. The issue of reliability and robustness of machines has been received a deep attention from researchers and practitioners maintenance. There has been an increased interest in machine condition monitoring because of the potential benefits to be obtained from reduced maintenance costs, improved operating efficiency, increased machine reliability an availability. Recently, the most fundamental issue of condition monitoring in industries are fault diagnosis and prognosis. One of the most effective to investigate in this issue is condition monitoring routine based on vibration signal analysis. However, current signal analysis can also be used in condition monitoring of electrical machine such as induction motors as well as vibration signal. Hence, the motivation of this research is to establish an intelligent condition monitoring, fault diagnosis and prognosis system which can be effectively applied in machines based on vibration and current signal analysis augmented by a kind of intelligent system method namely support vector machine (SVM).

The significance of this research is to develop the existed algorithm in SVM, so that it can perform well in machine fault diagnosis and prognosis. In this research, the developed system is addressed to be able to achieve good performance, high accuracy and robust in machine fault diagnosis routine using classification procedure.

SVM was selected technique to be applied to machine fault diagnosis process. The reason is that SVM has excellent ability in generalization process. In addition, classical learning approaches are designed to minimize error on the training dataset and it is called the empirical risk minimization (ERM). Those learning methods follow the ERM principle and neural networks are the most common example of ERM. On the other hand, the SVM are based on the structural risk minimization (SRM) principle rooted in the statistical learning theory. It gives better generalization abilities and SRM is achieved through a minimization of the upper bound of the generalization error [5-7].

3. Aims and Objectives

This research focuses on the development of existed method of SVM algorithm for machine fault diagnosis and prognosis. The aim of this research is to redevelop and modify SVM algorithm and to combine SVM algorithm with other cooperate method for obtaining the better performance in classification process using SVM. The main objectives of this research are as follows:

1. To redevelop preprocessing method of feature extraction and reduction for obtaining better SVM inputs by component analysis using linear and nonlinear technique.
2. To incorporate SVM with feature extraction and reduction using component analysis.
3. To redevelop a new kernel method using wavelet function and apply it to SVM based classifier.
4. To apply the developed system of SVM based-classifier in machine fault diagnosis.
5. To redevelop SVM based on regression for prognosis of machine condition.

4. Research Method and Approach

In order to achieve the aims and objectives of the research, the following quantitative research method has been adopted:

1. Theory redevelopment of component analysis that consists of multivariate data analysis using linear technique such as principal component analysis (PCA) and independent component analysis (ICA).
2. Redevelop the nonlinear technique of multivariate data analysis using kernel function and induce it in PCA and ICA.
3. Redevelop wavelet theory as kernel function for nonlinear mapping process in SVM.
4. Applying the redeveloped technique to induction motors fault diagnosis.
5. Study the feasibility of SVM based regression for prognosis of machine condition.

5. Contribution of This Research

The main contribution of this research is redeveloping the SVM algorithm for machine fault diagnosis and prognosis. Several other significant contributions of the redeveloped SVM algorithm technique are as follows:

1. The ability to obtain the optimal features for fault classification using feature extraction and reduction by linear and nonlinear technique of component analysis.
2. Establishing wavelet support vector machine (W-SVM) to gain a good performance and novelty in machine condition monitoring and fault diagnosis system.
3. The developed system was successfully applied in real application to diagnose and detect faults in induction motors based on vibration and

current signals.

4. Developing SVM based regression for prognosis of machine condition.

6. Organizational Overview of This Dissertation

Based on aforementioned aims and objectives of this research, this dissertation is outlined as follows.

Chapter 1 explains the background and motivation behind this research as well as the existed method and algorithm which be adopted through the appropriate research method. It also describes the main objectives and contributions of this research and outlines an overview of this dissertation.

Chapter 2 outlines the preliminary literature review and knowledge of fault diagnosis techniques, particularly in the time and frequency domain. Moreover, it discusses the feature-based fault diagnosis concept, statistical features representation, and data preprocessing.

Chapter 3 reviews the dimensionality reduction, concept of component analysis both using linear and nonlinear techniques. In addition, the basic theory of support vector machine (SVM) classifier and the frame work of building kernel function using wavelet for SVM classifier are deeply reviewed.

Chapter 4 considers to the real application on induction motor. It presents the faults frequently occurred in induction motor, diagnosis methods, the proposed diagnostic system and case study of fault diagnosis of induction motor using SVM incorporate with component analysis procedure. Moreover, it also presents fault diagnosis method using transient current signal analysis. It includes the preprocessing of transient current signal, statistical features representation, feature extraction and reduction and classification process using wavelet-support vector machine (W-SVM)

Chapter 5 addresses to use support vector regression for prognosis of machine condition.

Chapter 6 gives several conclusion based on the results obtained in this research. This chapter also recommends some directions for further research in the future.

References

[1] R.K. Mobley, An Introduction to Predictive Maintenance, Van Nostrand Reinhold, New York, 1990.

[2] E. Turban, J.E. Aronson, Decision Support Systems and Intelligent Systems, 6th Edition, Prentice International Hall, Hong Kong, 2001.

[3] J.H Williams, A. Davies, P.R. Drake, Condition-based Maintenance and Machine Diagnostics, Chapman & Hall, London, 1994.

[4] C.M. Harris, Shock and Vibration Handbook, McGraw-Hill, Singapore, 1997.

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

[6] B. Boser, I. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers, Proceedings of Fifth Annual Workshop on Computational Learning Theory, New York, 1992.

[7] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20(3) (1995) 273-297.

II. Preliminary Review and Study

1. Existing Method for Machine Condition Monitoring and Fault Diagnosis

1.1. Statistical Approach

A common method of fault diagnosis is to detect whether a specific fault is present or not based on the available condition monitoring without intrusive inspection of machine. In the early development method of fault diagnosis, a statistic test was constructed to summarize the condition monitoring information so as to be able to decide whether to accept or reject some hypothesis of machine condition [1-3]. Recently, a framework for fault diagnosis called structured hypothesis test was proposed for conveniently handling complicated multiple faults of different types [4].

Other fault detection and diagnosis technique was employed using statistical process control (SPC) which was originally developed in quality control theory. The principle of SPC is to measure the deviation of the current signal from a reference signal representing the normal condition to see whether the current signal is within the control limit or not. An example of using SPC for damage detection was discussed in [5].

Cluster analysis, as a multivariate statistical analysis method, is a statistical classification approach that groups signals into different fault categories on the basis of the similarity of the characteristics or features they possess. It seeks to minimize within-group variance and maximize between-group variance. The result of cluster analysis is a number of heterogeneous groups with homogeneous contents. There are substantial differences between the groups, but the signals within a single group are similar. Application of cluster analysis in machinery fault diagnosis was discussed in [6,7]. A natural way of signal grouping is based on certain distance measures or similarity measure between two signals. These measures are usually derived from certain discriminant functions in statistical pattern recognition [8]. Commonly used distance measures are Euclidean distance, Mahalanobis distance, Kullback–Leibler distance and Bayesian distance. Papers in [9-12] contain some examples of using these distance metrics for fault diagnosis. Ding et al. [9] introduced a new distance metric called quotient distance for engine fault diagnosis. Pan et al. [13] proposed an extended symmetric Itakura distance for signals in time–frequency representations such as the Wigner–Ville distributions. Other than distance measures, feature vector correlation coefficient is also a similarity measure commonly used for signal classification in machinery fault diagnosis [12]. Many clustering algorithms are available for determining the signal groups [14]. A commonly used algorithm in machine fault classification is the nearest neighbor algorithm that fuses two closest groups into a new group and calculates distance between two groups as the distance of the nearest neighbor in the two separate groups [15]. The boundary of two adjacent groups is determined by the discriminant function used. A piecewise linear discriminant function was used and thus piecewise linear boundaries were obtained for bearing condition classification in [16]. A technique called support vector machine (SVM) is usually employed to optimize a boundary curve in the sense that the distance of the closest point to the boundary curve is maximized. SVM applied to machine fault diagnosis was considered in [17,18].

1.2. Artificial Intelligent (AI) Approach

AI techniques have been increasingly applied to machine diagnosis and have shown improved performance over conventional approaches. In practice, however, it is not easy to apply AI techniques due to the lack of efficient procedures to obtain training data and specific knowledge, which are required to train the models. So far, most of the applications in the literature just used experimental data for model training. In the literature, two popular AI techniques for machine diagnosis are artificial neural networks (ANNs) and ESs. Other AI techniques used include fuzzy logic systems, fuzzy–neural networks (FNNs), neural–fuzzy systems and evolutionary algorithms (EAs). A review of recent developments in applications of AI techniques for induction machine stator fault diagnostics was given by Siddique et al. [19].

An ANN is a computational model that mimics the human brain structure. It consists of simple processing elements connected in a complex layer structure which enables the model to approximate a complex non-linear function with multi-input and multi-output. A processing element comprises a node and a weight. The ANN learns the unknown function by adjusting its weights with observations of input and output. This process is usually called training of an ANN. There are various neural network models. Feed-forward neural network (FFNN) structure is the most widely used neural network structure in machine fault diagnosis [20-23]. A special FFNN, multilayer perceptron with the BP training algorithm, is the most commonly used neural network model for pattern recognition and classification, and hence machine fault diagnostics as well [24,25]. The BP neural networks, however, have two main limitations: (1) difficulty of determining the network structure and the number of nodes; (2) slow convergence of the training process. A cascade correlation neural network (CCNN) does not require initial determination of the network structure and the number of nodes. CCNN can be used in cases where on-line training is preferable. Spoerre [26] applied CCNN to bearing fault classification and showed that CCNN can result in utilizing the minimum network structure for fault recognition with satisfied accuracy. Other neural network models applied in machine diagnostics are radial basis function neural networks, recurrent neural networks [27,28] and counter propagation neural networks [29]. The above ANN models usually use supervised learning algorithms which require external input such as the a priori knowledge about the target or desired output. For example, a common practice of training a neural network model is to use a set of experimental data with known (seeded) faults. This training process is supervised learning. In contrast to supervised learning, unsupervised learning does not require external input. An unsupervised neural network learns itself using new information available.

Wang and Too [30] applied the unsupervised neural networks, self-organizing map (SOM) and learning vector quantization to rotating machine fault detection. Tallam et al. [31] proposed some self-commissioning and on-line training algorithms for FFNN with particular application to electric machine fault diagnostics. Sohn et al. [3] used an auto associative neural network to separate the effect of damage on the extracted features from those caused by the environmental and vibration variations of the system. Then a sequential probability ratio test was performed on the normalized features for damage classification. In contrast to neural networks, which learn knowledge by training on observed data with known inputs and outputs, ESs utilize domain expert knowledge in a computer program with an automated inference engine to perform reasoning for problem solving. Three main reasoning methods for ES used in the area of machinery diagnostics are rule-based reasoning [32-34], case-based reasoning [35,36] and model-based reasoning [37]. Another reasoning method, negative reasoning, was introduced to mechanical diagnosis by Hall et al. [38]. Stanek et al. [39] compared case-based and model-based reasoning and proposed to combine them for a lower-cost solution to machine condition assessment and diagnosis. Unlike other reasoning methods, negative reasoning deals with negative information, which by its absence or lack of symptoms is indicative of meaningful inferences. ESs and neural networks have their own limitations. One main limitation of rule-based ESs is combinatorial explosion, which refers to the computation problem caused when the number rule increases exponentially as the number of variables increases.

Another main limitation is consistency maintenance, which refers to the process by which the system decides when some of the variables need to be recomputed in response to changes.

2. SVM in Machine Condition Monitoring and Fault Diagnosis: a review

Recently, the issue of machine condition monitoring and fault diagnosis as a part of maintenance system became global due to the potential advantages to be gained from reduced maintenance costs, improved productivity and increased machine availability. This sub chapter presents a survey of machine condition monitoring and fault diagnosis using support vector machine (SVM). It attempts to summarize and review the recent research and development of SVM in machine condition monitoring and fault diagnosis. Numerous methods have been developed based on intelligent system such as artificial neural network, fuzzy expert system, condition-based reasoning, random forest, etc. However, the use of SVM for machine condition monitoring and fault diagnosis is still rare. SVM has an excellence performance in generalization, so it can produce high accuracy in classification for machine condition monitoring and diagnosis. Until 2006, the use of SVM in machine condition monitoring and fault diagnosis is tending to develop towards expertise orientation and problem-oriented domain. Therefore, the ability to continually change and obtain a new novel idea for machine condition monitoring and fault diagnosis using SVM will be a future works.

2.1. Diagnosis of Rolling Element Bearing

Bearings are the best location for measuring machinery vibration since this is where the basic dynamic loads and forces of machine are applied and they are a critical component of machinery. Condition monitoring and fault diagnosis of bearing can represent the condition of machine itself. This section will review the authors who have contribution in research of fault diagnosis of bearing.

Jack and Nandi [40] performed fault detection of roller bearing using SVM and artificial neural network (ANN). They used vibration data taken from small test rig and simulate the bearing condition which has four faults: inner race fault, outer race fault, rolling element fault and cage fault. They defined and calculated statistical features based on moments and cumulants and selected the optimal features using GA. In the classification process, they employed SVM using RBF kernel with constant kernel parameter. Yan and Shao [41] employed SVM for fault detection of roller bearing using vibration signal and noise. Unfortunately, there is no special method stated in their research except SVM classification routine. However, they stated that SVM has promising application in fault diagnosis. Moreover, Samanta et al. [42, 43] have improved the previous methods in fault detection of bearing. They applied GA for feature selection and searching proper RBF kernel parameters. Several effect conditions such as sensor location, signal preprocessing, number of features were presented to show the performance of SVM compared with ANN. Rojas and Nandi [44, 45] have improved their previous research on bearing fault diagnosis. They proposed a practical scheme for fast detection and classification of rolling element bearing. Sequential minimal optimization (SMO) was implemented for solving SVM optimization problem. Zhang et al. [46] proposed probabilistic SVM (ProSVM) for fault diagnosis of bearing. It was aimed to effectively reduce the number of samples on the condition of keeping the classification accuracy. Sugumaran et al. [47] employed fault diagnosis of roller bearing using decision tree (DT) and proximal SVM (PSVM). DT was aimed to identify the best features from a given set of samples for the purpose of classification. They claimed that PSVM has the capability to efficiently classify the faults using statistical features. Recently, Hu et al. [48] proposed a method that used improved wavelet package transform (IWPT) and SVM ensemble for fault diagnosis of rolling element bearing. They also employed feature selection using distance evaluation technique (DET) for feature selection.

The previous discussion describes the evolutionary by years the technical development of using SVM for bearing fault diagnosis. Hopefully, there will be present an advanced research to obtain more robust techniques in SVM for bearing fault diagnosis.

2.2. Diagnosis of Induction Motors

Induction motor is a critical component in many industrial processes and it is very important part to support the survival of industry in producing of products. It is also frequently integrated with any commercially available equipment and the process itself. Therefore, it has been urgently required special attention in condition monitoring to guarantee the performance of induction motors. Early fault diagnosis of induction motor during its operation will give the incipient faults condition and the efforts to overcome any faults should be done to avoid the more serious condition.

Pöyhönen et al. [49, 50] proposed method namely coupling pairwise SVM for fault classification of induction motors. Power estimate density using Welch's method was calculated from circulating currents in parallel branches of motors. SVM was then trained to distinguish a healthy spectrum from faulty spectra. The induction motors consist of faults as follows: broken rotor bars, broken end-ring in rotor cage, shorted coil and shorted turn in stator winding. Zhitong et al. [51] carried out fault detection in induction motors using SVM technique to detect broken rotor bars. In their experiment, induction motors were experimented with no fault, one broken bar, two broken bars and three broken bars. They used stator current to obtain the signal and calculated the frequency spectrum for doing fault detection. Fang [52] conducted a faults diagnosis system based on integration of rough set theory (RST) and SVM. He used stator current spectrum as inputs. RST can perform feature extraction and reduction for removing redundant attributes.

The conditions of induction motors were health, broken bar, dynamic eccentricity and static eccentricity. The result showed that the proposed method has good performance in diagnosis accuracy and needs short time in training. Widodo and Yang [53-55] employed fault diagnosis method using SVM combined by feature extraction via component analysis (PCA, ICA, KPCA and KICA). The statistical features in time domain and frequency domain from current and vibration signal were calculated as features representation. The proposed method was aimed to detect fault in induction motor such as broken rotor bars, bowed rotor, bearing fault, rotor unbalance, eccentricity and phase unbalance. Recently, they conducted fault diagnosis of induction motor based on start-up transient current signal. Transient current signal has characteristic (similarity) that was difficult to distinguish among faults. Therefore, they proposed wavelet SVM (W-SVM) for obtain a novel method in classification process. The basic idea of W-SVM was constructing a kernel function using wavelet function and then inducing into SVM theory [56, 57].

2.3. Diagnosis of Machine Tools

Recently, AI technique has been used for fault detection of machine tool. Moreover, AI can also predict the remaining life of machine tools. Here, the survey of using SVM for condition monitoring and diagnosis of machine tools will be presented.

Ramesh et al. [58] presented a hybrid SVM-Bayesian Network (BN) for predicting the thermal error in machine tool according to specific condition. In this research, SVM-BN was developed first all to classify the error into groups depending on the operating condition and then carry out a mapping of the temperature profile with the measured error. This concept lead to a more generalized prediction model then the conventional method of directly mapping error and temperature irrespective of condition. Such model is especially useful in a production environment wherein the machine tools are subject to a variety of operating conditions. The other research was carried out by Sun et al. [59, 60] who classified tool wear using SVM based on manufacturing consideration. This research was aimed to propose a new performance evaluation function for tool condition monitoring (TCM). First, they analyzed two types of manufacturing loss due to misclassification (loss caused under prediction and over prediction) then both are utilized to compute corresponding weights of the proposed performance evaluation function. Then the expected loss of future misclassification is introduced to evaluate the recognition performance of TCM. Finally, a revised SVM approach is implemented to carry out the multi-classification of tool states. With this approach, a tool is replaced or continued not only based on the tool condition alone, but also the risk in cost incurred due to underutilized or overused tool. In recent publication, Cho [61] conducted TCM for tool breakage detection using SVM in milling process. SVM was addressed to recognize process abnormalities and initiate corrective action during a manufacturing process. They applied support vector regression (SVR) for tool breakage determination and claimed better than traditional multiple variable regression approach (MVR).

The survey of papers which implement SVM in TCM has been presented. However, there are only few paper discuss about TCM during year 1999-2006 according to survey from some on-line journals.

2.4. Diagnosis of Pumps, Compressors and Turbines

Detection of pump failure has been carried out by Tax et al. [62] using support vector data description (SVDD). The importance of preprocessing data was also highlighted in this dissertation such as feature extraction and selection. In addition, they evaluated several feature extraction methods in a special type of outlier detection problem. The use of support vector data description was aimed to get indication of the complexity of the normal class in data sets and how well it is expected to be distinguishable from the abnormal data. Gao et al. [63] applied SVM for fault diagnosis of valve in reciprocating pumps. As preprocessing, the wavelet packet transform (WPT) was employed to extract feature vectors from vibration signal. They simulate 10 conditions of valve which must be classified using SVM. SVM was successful applied for fault diagnosis of turbo-pump rotor by Yuan et al. [64]. The original data came from vibration signal then the feature extraction was performed by applying principal component analysis (PCA) to extract the optimal features and reduce the dimension of features. In addition, based on same data, Yuan [65] was also carried out fault diagnosis of turbo-pump using SVM with parameter optimization. In this research, artificial immunization algorithm (AIA) was used to optimize parameters in SVM.

Yang et al. [66] performed condition classification of small reciprocating compressor for refrigerator using SVM. In this dissertation, wavelet transform and statistical method were used to extract salient features from row noise and vibration signal. Moreover, iteration method was employed to select the proper RBF kernel parameters in SVM. In addition, Yang et al. [67] also carried out cavitation detection of butterfly valve using SVM. The other research using SVM for fault diagnosis of reciprocating compressor was performed by Ren et al. [68]. This was aimed to detect valve fault using vibration signal. Vibration signal was decomposed using local wave method and data was acquired from valve surface.

In turbine detection, Li et al. [69] employed SVM for fault diagnosis of steam turbine. Ensemble learning based on genetic algorithm (GA), namely direct genetic ensemble (DGE) was performed to achieve good performance in classification. The proposed system successfully detected rotor unbalance in steam turbine. Zhang et al. [70] successfully applied fuzzy support vector machine (FSVM) for condition monitoring of flue-gas turbine set based on vibration signal. FSVM modified separating hyperplane by adding fuzzy coefficients to every training data sample in order to indicate loss difference of misclassifying training data sample of different types.

2.5. Diagnosis of HVAC Machines

Batur et al. [71] presented fault detection of heat exchanger using SVM combined by least squares parameter identification technique to permit on-line monitoring. In this system, SVM was addressed to detect abnormal condition of heat exchanger such as air in steam line, obstructed tubes, high condensate flow and low condensate flow. The other research was conducted by Han et al. [72] for hot spot detection in power plant boiler air preheater based on least squares support vector machine (LS-SVM). In this system, discriminate model of 3 pairs of fire status have been built based on LS-SVM using RBF kernel and polynomial kernel. The hyperparameters of classifiers were tuned by leave-one-out cross validation. Receiver operating characteristic (ROC) curve showed that LS-SVM has good classification and generalization ability. Choi et al. [73] carried out fault diagnosis of chillers machine using SVM based on statistical test such as generalized likelihood ratio (GLR). The system was subjected to five types of faults, including reduction in water flow rates in condenser, evaporator, fouling in condenser and evaporator and refrigerant undercharge.

2.6. Other Machines

The other applications of using SVM for machine condition monitoring and fault diagnosis are reported as follows: Rychetsky et al. [74] employed support vector machine for engine knock detection. In this research, support vector was combined by kernel adatron technique to provide non linearity, a bias and soft margin. This kernel adatron algorithm was reported can be convergence fast and proper for combination with SVM. SVM classifier was addressed to classify the current knocking condition (3 classes): no-knock, borderline knock and hard knocking. Xu et al. [75] employed rough set theory combined with SVM (RST-SVM) for fault detection of diesel engine. Fault diagnosis of diesel engines is difficult problem due to the complex structure of the engine and the presence of multi-excite sources. In this dissertation, diagnosis procedure was addressed to diagnose fault conditions such as intake clearance is too small, intake valve clearance is too large and exhaust valve clearance is too large. Moreover, integrating the advantages of RST in effectively deal with the uncertainty information and SVM produced greater generalization performance. The diagnosis of the diesel demonstrated that the solution can reduce the cost and raise the efficiency of diagnosis, and verified the feasibility of engineering application. Hu et al. [76] developed method called fusion multi-class SVM for fault diagnosis of diesel engine. The main idea of this method is combining the information of several sources then constructs it as an input space. Then SVM classifier realized classification by finding the optimal hyperplane with a maximal margin. The system used vibration signal from three accelerometers which attached on first cylinder head, second cylinder head and the center of the piston stroke. Four conditions were simulated for diagnosing process those are intake valve clearance is too small; intake valve clearance is too large; exhaust valve clearance is too large and normal condition. The result showed that the proposed method can largely improve the diagnosis accuracy.

In addition, SVM was also reported in application of fault diagnosis in sheet metal stamping operation. The research was conducted by Ge at al. [77], they used strain signal of stamping process which are highly nonlinear and non-stationary and it was typical signal in metal forming process. In this experiment, two kinds of operation of metal stamping were tested, the first one was a single step blanking and the second one was a multi-step progressive stamping. The conditions for simulating the process were no fault, misfeed (work piece is not aligned with the dies), slug (the leftover of the position hole is left on the surface of the upcoming work piece), work piece is too thick, work piece is too thin and work piece material is missing.

Samanta [42] carried out gear fault detection using SVM combined with genetic algorithm. The time domain vibration signal of a rotating machine with normal and defective gears are processed for feature extraction. The extracted features from original signal were used as inputs to SVM classifier. In this research, GA was performed in feature selection and optimizing RBF kernel parameters.

Aforementioned survey gives the description of application of SVM in machine condition monitoring and diagnosis. Actually, it can be said that only few papers found in this application rather than other areas such as described in previous chapter.

3. Feature-Based Diagnosis Concept

The process of traditional condition monitoring and fault diagnosis can be summarized as follows: data acquisition, data preprocessing, data analysis and decision making. Here, the data that represents the machine condition called condition-based monitoring. Nevertheless, there is problem of such a system in data transferring and storage. For instance, when monitoring of large system of rotating machinery will be performed, the installation of many sensors is needed to assure the diagnosis reliability. Such many sensors result in huge dimensionality of data.

With the globalization and fast growth of the computer and information technology, on-line condition monitoring and fault diagnosis has gained much attention. Data transfer and storage problem become serious. If direct transferring a plenty of raw data will be performed so long time delays due to heavy traffic may be experienced which results in the lost of monitoring and diagnosis.

Therefore, representing data as features is a best solution for this problem that is expected able greatly reduces the requirement of transfer number and save storage space. The represented data as feature is similar to compress the data from many domains with keeping the information as high as possible. Thus, a relative technique has came out such as feature representation, feature extraction and feature selection.

The typical feature-based condition monitoring and diagnosis framework is illustrated in Fig. 2.1, which can be summarized as follows:

- The data are on-line acquired from object machine as a raw data that need preprocessing to condition the data as good as possible for emerging the salient condition of machine. These data can be vibration signals, current and volt signals, sound signals, flux signals, etc. Corresponding to object, the different preprocessing procedure can be used such as filtering (high, low and band-pass), wavelet transform, averaging. Smoothing and so on.
- The features are calculated from various domains: time, frequency, cepstrum or wavelet domain. In this way, the information of raw data is kept as good as possible to address the analysis method in the next. Furthermore, the transfer and storage problem of data can be solved.
- Many calculations of feature parameters in many domains result high dimensionality of data features. All of them are not useful for condition analysis; sometimes some of them even can increase the difficulty of analysis and degrade the accuracy. So reducing the dimension of data features is necessary which can remove the irrelative and garbage features.
- According to monitoring object, the features which can significantly represent machine performance should be selected.
- The selected features are then sent to condition monitoring and fault diagnosis system to define the machine condition.

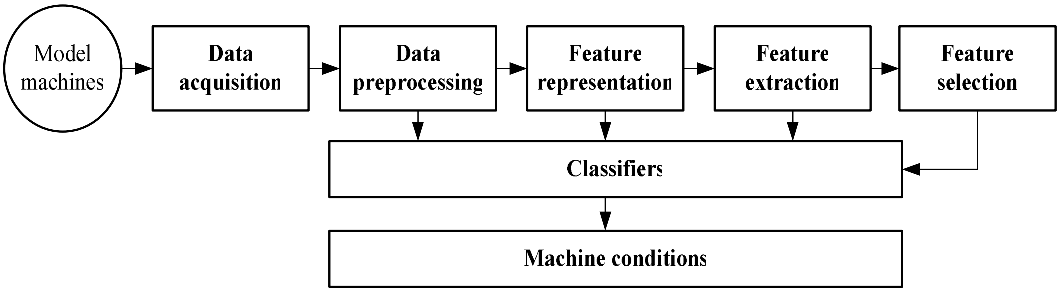


Fig. 2.1 Feature-based condition monitoring and fault diagnosis system [78].

Usually, condition monitoring and fault diagnosis system works based on pattern classification. For feature-based diagnosis system, the quality of data receives deep attention to guarantee the accuracy of diagnosis process. Therefore, the preprocessing of data is important step. A good preprocessing will reduce the noise in the data and retains as much information as possible [79]. When the number of objects in the training set is too small for the number of feature used, most of classification procedures cannot find good classification boundaries. This is called curse of dimensionality [80]. By a good preprocessing the number of feature per object can be reduced such that the classification problem can be solved well.

Feature-based diagnostic procedure has been employed for fault diagnosis of machine. In this case, it needs feature extraction procedure which is addressed to obtain the optimal features from original data set. Tax et al. [62] employed feature-based procedure from power spectrum, envelope spectrum, autoregressive modeling, music spectrum and classical spectrum for failure detection of a small submersible pump. They tried to find the best representation of data features such that the target class can best be distinguished from the outlier class. The support vector data description was proposed to accomplish their work for finding the smallest sphere containing all target data.

The authors who used statistical features based on moments, cumulants and other statistical features of the time data series and spectral of vibration data for fault detection are reported in [41-44]. Yang et al. [66, 67, 81-84] used statistical features of time domain and frequency domain for fault detection in rotating machinery and cavitations in butterfly valve. In the case of induction motor, they acquired data of vibration and stator current signal. Yuan et al. [64, 65] performed fault diagnosis of turbo-pump rotor using data features which is acquired from frequency bands of secondary vibration signal. The frequency of secondary signal is divided into 9 bands then the frequency amplitudes on each band and their average value are calculated as features. Sun et al. [59, 60] employed statistical features which came from acoustic emission signal for detection wear in machine tool. They also used cutting parameter such as cutting speed, depth of cut and feed rate as additional features. Cho et al. [61] carried out tool break detection using features from cutting forces and power consumption in end milling machine. The other application was reported that Han et al. [72] conducted hot spot detection in power plant using features from data temperature which acquired by thermocouple and infrared sensors. Moreover, Ramesh et al. [58] conducted a prediction of thermal error in machine tools using features from temperature sensors.

In feature-based diagnosis process, after defining the features i.e. statistical features from original data acquisition, the huge dimensionality problem of data features is possibly emerged. It cannot be avoided because of not all features are useful and optimal for classification process. The existence of irrelative features tends to degrade the performance of classifier. One of solutions to solve this problem is performing feature extraction which can extract the optimal features and all at once reduce the dimensionality of features. Basically, feature extraction means mapping process of data from higher dimension into low dimension space. Many methods have been proposed to perform dimensionality reduction using linear and nonlinear techniques. In machine condition monitoring and fault diagnosis research area, feature extraction using component analysis was reported as follows: using linear method, principal component analysis (PCA) was conducted by Widodo and Yang [53, 54] and Yuan et al. [65], using independent component analysis (ICA) [53, 54]. Moreover, nonlinear feature extraction using kernel PCA and kernel ICA was also performed by Widodo and Yang [55]. The other techniques called rough set theory was conducted for extracting optimal features and reduce dimension of features by Xu and Fang [52, 69]. In their research, rough set theory (RST) was employed to preprocess the data for eliminating redundant information and reducing the sample dimension.

The other hand, some researches suggested using feature selection after defining features set from original data. The techniques which are addressed to feature selection were genetic algorithm (GA) and distance evaluation technique (DET). In machine fault diagnostics area, researchers who employed GA technique were Jack and Nandi [40], Samanta et al. [42, 43], and Li et al. [69]. In addition, DET was reported successfully doing feature selection by Yang et al. [81, 83] and Hu et al. [48].

From aforementioned discussion, it can be observed that feature-based diagnosis has been widely used in many applications of machine condition monitoring and fault diagnosis. Most of results of feature-based technique were relatively satisfied according to previous papers. It means that feature-based procedure is strongly suggested when recognition and classification process will be performed.

4. Statistical Feature Representation

Usually, in the application of pattern classification and recognition, the data are represented by features which can be characteristic values, colors and so on. In machine condition monitoring and fault diagnosis, the statistical features are selected as patterns which can indicate the machine condition. Furthermore, statistical features are simple and useful for exploring and indicating the incipient faults when faults occurred in machines.

This section focuses on feature representation of statistical features for machine condition monitoring and fault diagnosis. Transformation of data to features plays a very important role which directly affects the performance of whole system. In other words, the better the features can reflect the performance of task the better the result will be. In order to keep data information at the highest level, features are calculated from the time domain, frequency domain and auto-regression estimation.

4.1. Features in Time Domain

4.1.1. Cumulants

The features described here are called statistical features because they are based on only the distribution of signal samples with the time series treated as a random variable. These features were also known as moments or cumulants. In most cases, the probability density function (pdf) can be decomposed into its constituent moments. If a change in condition causes a change in the probability density function of the signal then the moments may also change. Therefore, monitoring this phenomenon can provide diagnostic information.

The moment coefficients of time waveform data can be calculated using following equations

$$m_n = E\{x^n\} = \frac{1}{N}\sum_{i=1}^{N} x_i^n \tag{2.1}$$

where $E\{\cdot\}$ represents the expected value of the function, x_i is the ith time historical data and N is the number of data points.

The first four cumulants: mean (c_1), standard deviation (c_2), skewness (c_3) and kurtosis (c_4), can be calculated from the first four moments using the following relationships

$$c_1 = m_1 \tag{2.2}$$

$$c_2 = m_2 - m_1^2 \tag{2.3}$$

$$c_3 = m_3 - 3m_2m_1 + 2m_1^3 \tag{2.4}$$

$$c_4 = m_4 - 3m_2^3 - 4m_3m_1 + 12m_2m_1^2 - 6m_1^4 \tag{2.5}$$

In addition, non-dimensional feature parameters in time domain are more popular such as shape factor and crest factor.

$$SF = x_{rms} / x_{abs} \tag{2.6}$$

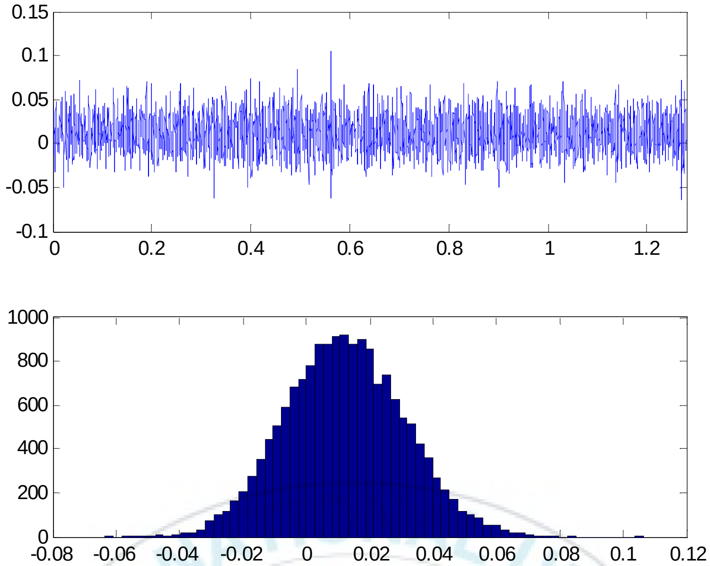
$$CF = x_p / x_{rms} \tag{2.7}$$

where x_{rms}, x_{abs} and x_p are root mean square value, absolute value and peak value, respectively.

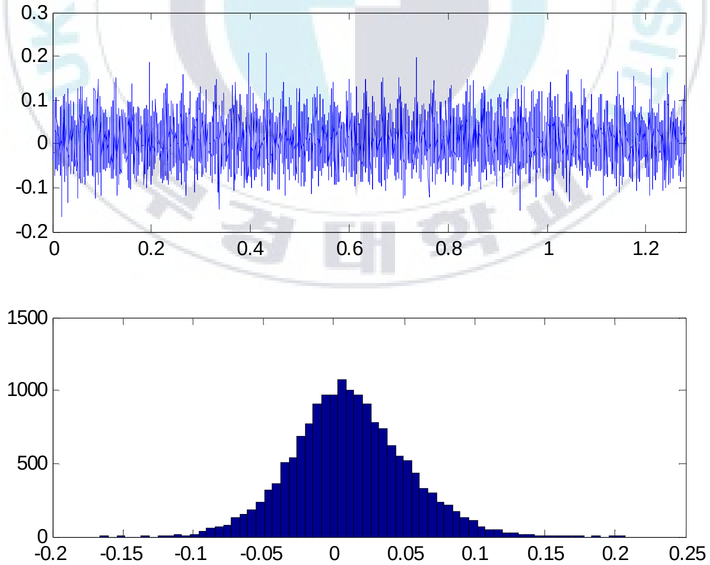
Fig. 2.2 describes the bearing signals and its histogram with different condition (normal and faults). Moreover, the cumulants are highlighted according to bearing condition and its values are summarized in Table 2.1.

Table 2.1 Cumulants for bearing signal with different condition

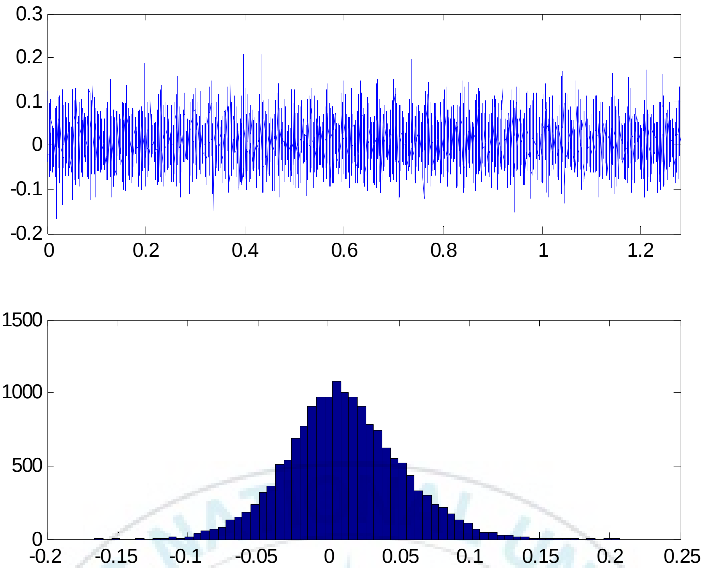
Cumulants	Conditions			
	Normal	Unbalance	Inner-race	Misalignment
Mean	0.0122	0.085	0.0038	0.0507
STD	0.0188	0.0314	0.0821	0.1833
Skewness	0.0802	0.0184	0.1234	0.1597
Kurtosis	3.0332	3.282	7.083	3.4315



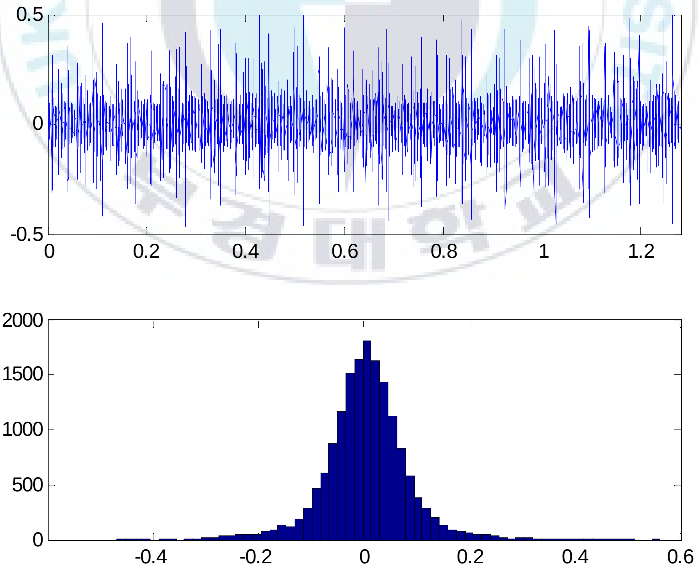
(a) Normal condition



(b) Unbalane



(c) Misalignment



(d) Inner race fault

Fig. 2.2 Histogram of bearing signal with different conditions.

From Table 2.1, kurtosis has clear information for describing the condition of bearing. Normal bearing has kurtosis 3.0 while the fault condition has kurtosis more than 3.0. Therefore, kurtosis can indicate the incipient fault at bearing during its operation.

4.1.2. Histogram: Upper and Lower Bound

Histograms which can be thought as a discrete probability density function (pdf) are calculated in the following way. Let *d* be the number of divisions that need to divide range into, let h_i with $0 \le i \le d$ be the columns of the histogram, then

$$h_i = \sum_{j=0}^{n} \frac{1}{n} r_i(x_i), \forall i, 0 \le i \le d \tag{2.8}$$

$$r_i(x) = \begin{cases} 1, & \text{if } \dfrac{i(\max(x_i) - \min(x_i))}{d} \le x < \dfrac{(i+1)(\max(x_i) - \min(x_i))}{d} \\ 0, & otherwise \end{cases} \tag{2.9}$$

The upper bound and lower bound of histogram are defined as

$$h_L = \max(x_i) - \Delta / 2 \tag{2.10}$$

$$h_U = \max(x_i) + \Delta / 2 \tag{2.11}$$

where $\Delta = \max(x_i) - \min(x_i) / (n-1)$

Effectively, it is normalized by two things: the length of the sequence and the column divisions. Since the term above includes a $1/n$ term and every x_i must fall into exactly one h_i column, the next effect is that h_i = 1 (i = 0,…, d-1). The column divisions are relative to the bounding box and thus most of h_i will not be zero. This is a desirable, since it essentially removes the issue of size of a sign, and low resolution on small signs with lots of empty columns. The alternative would be to have absolute locations which would be nowhere near as closely correlated with the information in the sign itself. The examples of histogram of bearing signal with different condition can be seen in Fig. 2.2.

4.1.3. Entropy Estimation and Error

In information theory, uncertainty can be measured by entropy. The entropy of distribution is the amount of a randomness of the distribution. Entropy estimation is two stage processes; first, a histogram is estimated and thereafter the entropy is calculated. The entropy estimation $E_s(x_i)$ and standard error $E_e(x_i)$ are defined as

$$E_s(x_i) = -\sum P(x_i)\ln(Px_i) \tag{2.12}$$

$$E_e(x_i) = \sum P(x_i)\ln P(x_i)^2 \tag{2.13}$$

Where x_i is discrete time signals, $P(x_i)$ is the distribution on the whole signal. Here, the entropy of vibration and current signals are estimated using unbiased estimated approach.

4.2. Features in Frequency Domain

Through the frequency domain parameter indices, the primary diagnosis is available. In other words, the features can indicate the faults. Because these calculation indices are simple and fast so they can be used in the on-line condition monitoring. When there are some changes on the parameters, it indicates occurrence of faults. Finally, the precision diagnosis can deal with the problem.

For example, the signal of ball bearing are composed of many stochastic elements, different faults have different spectrum in the frequency domain. However, in some cases the faults cannot be distinguished by power spectrum. Above mentioned, frequency parameters indices can show the faults in the beginning of the failure. So these indices can be used to perform condition monitoring and fault diagnosis.

The signal power spectrum shows the power distribution with the frequency. When the frequency elements and their power changed so the position of the main spectrum changed. On the other hand, when the frequency elements increased the power spectrum distribution become discrete whereas the power distribution is shown change. The characteristics of the frequency domain can be shown well through frequency parameter indices as follows:

Frequency center (FC)

$$FC = \frac{\int_0^{+\infty} fs(f)df}{\int_0^{+\infty} s(f)} \tag{2.14}$$

Mean square frequency (MSF)

$$MSF = \frac{\int_0^{+\infty} f^2 s(f)df}{\int_0^{+\infty} s(f)} \tag{2.15}$$

Root mean square frequency (RMSF)

$$RMSF = \sqrt{MSF} \tag{2.16}$$

Variance frequency (VF)

$$VF = \frac{\int_0^{+\infty} (f - FC)^2 s(f)df}{\int_0^{+\infty} s(f)df} \tag{2.17}$$

Root variance frequency (RVF)

$$RVF = \sqrt{VF} \tag{2.18}$$

where *s*(*f*) is the signal power spectrum. The *FC*, *MSF* and *RMSF* show the change position of main frequencies, *VF* and *RVF* describe the convergence of the spectrum power.

From the view of physics, the power spectrum is considered as the mass density function of a stick in the ordinate axis. Accordingly, *FC* is the mass center in the abscissa. When larger the density is near the origin, the distance between

FC and the origin is closer. *RMSF* is rotating radial circling the stick. The relation of the distance and density is same with aforementioned *FC*.

Due to real calculation, the frequency spectrum needs to be discrete

$$FC = \frac{\sum_{i=2}^{N} \dot{x}_i x_i}{2\pi \sum_{i=1}^{N} x_i^2} \tag{2.19}$$

$$MSF = \frac{\sum_{i=2}^{N} \dot{x}_i^2}{4\pi^2 \sum_{i=1}^{N} x_i^2} \tag{2.20}$$

$$VF = MSF - (FC)^2 \tag{2.21}$$

where $\dot{x}_i = x_i - \frac{x_{i-1}}{\Delta}$.

4.3. Auto Regression (AR) Coefficient

Since the different faults display different characteristics in the time series, auto-regression model is used to establish a model. The autoregressive coefficients are extracted as feature of fault condition. The first eight order coefficient of AR models are selected through Burg's lattice-based method using harmonic mean of forward and backward squared prediction errors. The definition that will be used here is as follows

$$y_t = \sum_{i=1}^{N} a_i y_{t-i} + \varepsilon_t \tag{2.22}$$

where a_i is the AR coefficients, y_t is the series under investigation, and N is the number of the model (N=8). The noise tem or residual ε_t is almost assumed to be Gaussian white noise.

5. Data Preprocessing

Data preprocessing describes any kinds of preprocessing on a raw data to enhance the reliability and thereby to improve the accuracy for signal analysis purpose. Data preprocessing may be performed on the data for understanding the nature of the data and extracting more meaningful knowledge from a given set of data. After data acquisition process, the problems with data often can be avoided. Several data problems can exist such as corrupt and noisy, irrelevant, missing attributes and so on. Therefore, data preprocessing is needed to enhance the quality of data for specific purposes i.e. pattern recognition and classification.

Data preprocessing transforms the data into a format that will be more easily and effectively to be processed appropriate with desire of user. As general, data preprocessing technique can be classified as follows:

- Transformation such as data filtering, data ordering, data editing, noise modeling, etc.
- Information gathering using data visualization, data elimination, data selection, sampling and so on.
- Generation of new information including adding new features, data fusion, data simulation, dimensional analysis, etc.

In followed section, data preprocessing is applied for obtaining the meaningful knowledge from raw data using wavelet transform, averaging, enveloping and cepstrum.

5.1. Wavelet Transform

The wavelet transform decomposes a concerned signal into a linear combination of time scale unit. It analyzes original signals and organizes them into several signal components according to the translation of the mother wavelet or wavelet basis function which changes the scale and show the transition of each frequency component.

5.1.1. Continuous wavelet transform (CWT)

The continuous wavelet transform is an integration with respect to the total time of the product of the target signal $f(t)$ and the mother wavelet $\psi_{a,b}$. Using mathematical expression, the continuous wavelet transform of the time function $f(t)$ca be written as

$$CWT(a,b) = \int_{-\infty}^{\infty} f(t)\psi_{a,b} dt \tag{2.23}$$

$$\psi_{a,b} = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{2.24}$$

where a,b and $\psi_{a,b}$ are the scale, translation parameters and mother wavelet, respectively.

The transform result represents the correlation between the signal and the transform of the mother wavelet being scaled and translated. If the signal and the mother wavelet are similar, the transform result will have a large value. This means that lead and delay are translation, while the scale is an expansion and compression. As the low scale is a compressing wavelet, it becomes a rapid changing signal, that is, it improves the sensitivity in the time domain for high frequency signals and improves the sensitivity in frequency domain for low frequency signals. This makes it possible to perform a multi-resolution analysis.

5.1.2. Discrete wavelet transform (DWT)

The orthogonal basis functions used in wavelet analysis are families of scaling function, $\phi(t)$ and associated wavelet $\psi(t)$. The scaling function can be represented by following mathematical expression

$$\phi_{j,k}(t) = \sum_{k} H_k \ \phi(2^j t - k) \tag{2.25}$$

where H_k represents coefficient of scaling function, k, j represent translation and scale, respectively.

Similarly, the associated wavelet can be generated using the same coefficient as the scaling function

$$\psi_{j,k}(t) = \sum_k (-1)^k \sqrt{2} h_{1-k}\ \phi(2^j t - k) \tag{2.26}$$

The scaling function is orthogonal to each other as well as with the wavelet function as shown in Eqs. (2.25) and (2.26). This fact is crucial and forms part of the framework for multi-resolution analysis.

$$\int_{-\infty}^{\infty} \phi(2k - t)\ \phi(2k - 1)\ dt = 0 \tag{2.27}$$

$$\int_{-\infty}^{\infty} \psi(t)\ \phi(t)\ dt = 0 \tag{2.28}$$

Using an iterative method, the scaling function and associated wavelet can be computed if the coefficients are known. Fig. 2.3 shows the Daubechies 2 and 5 scaling function and wavelet.

A signal can be decomposed into approximate coefficients $a_{j,k}$ through the inner product of the original signal at scale j and the scaling function.

$$a_{j,k} = \int_{-\infty}^{\infty} f_j(t) \phi_{j,k}(t) dt \tag{2.29}$$

$$\phi_{j,k}(t) = 2^{-j/2} \phi(2^{-j} t - k) \tag{2.30}$$

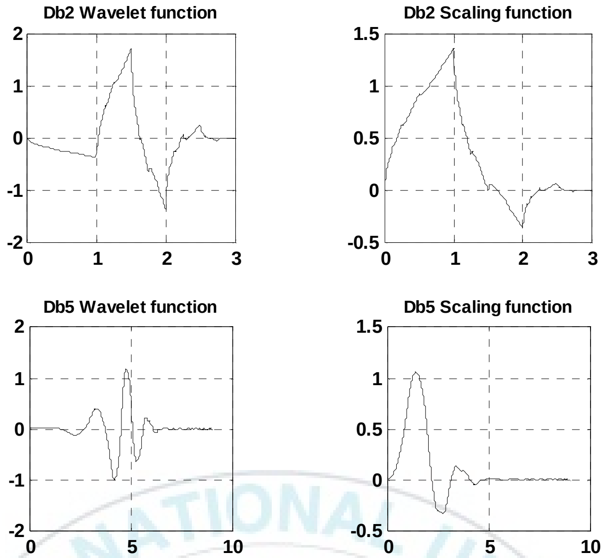


Fig. 2.3 Daubechies 2 and 5 scaling function and associated wavelet.

Similarly the detail coefficients $d_{j,k}$ can be obtained through the inner product of the signal and the complex conjugate of the wavelet function.

$$d_{j,k} = \int_{-\infty}^{\infty} f_j(t)\, \psi_{j,k}(t)\, dt \tag{2.31}$$

$$\psi_{j,k}(t) = 2^{-j/2}\phi(2^{-j}t - k) \tag{2.32}$$

The original signal can therefore be decomposed at different scales as follows

$$f(t) = \sum_{j=-\infty}^{\infty} a_{j_0,k}\ \phi_{j_0,k} + \sum_{j=-\infty}^{j_0}\sum_{k=-\infty}^{\infty} d_{j,k}\ \psi_{j,k}(t) \tag{2.33}$$

$$f[n] = \sum_{k=0}^{N-1} a_{j,k}\ \phi_{j,k}(t) = \sum_{k=0}^{N-1} a_{(j+1),k}\ \phi_{(j+1),k}(t) + \sum_{k=0}^{N-1} d_{(j+1),k}\ \psi_{(j+1),k}(t) \tag{2.34}$$

The coefficient of the next decomposition level ($j+1$) can be expressed as

$$a_{(j+1),k} = \sum_{k=0}^{N} a_{j,k} \int \phi_{j,k(t)}\ \phi_{(j+1),k}(t)\ dt \tag{2.35}$$

$$d_{(j+1),k} = \sum_{k=0}^{N} a_{j,k} \int \phi_{j,k(t)}\ \psi_{(j+1),k}(t)\ dt \tag{2.36}$$

$$a_{(j+1),k} = \sum_k a_{j,k}\ g[k] \text{ and } d_{(j+1),k} = \sum_k a_{j,k}\ h[k] \tag{2.37}$$

The decomposition coefficients can therefore be determined through convolution and implemented by using a filter. The filter $g[k]$ is a low-pass filter and $h[k]$ is a high-pass filter.

$$y[n] = \sum_{k=1}^{N} h[k]\ x[n-k] \tag{2.38}$$

5.2. Averaging

Averaging can be divided into two types: one is synchronous averaging and the other is spectrum averaging. Synchronous averaging is very useful in reducing the random noise component in the measurement or in reducing the effect of the other interfering signals such as components from another nearby machine which requires a tachometer to synchronize each snapshot of the signal to the running speed of machine. Unlike synchronous averaging, spectrum averaging does not reduce the noise. Instead, it finds the average magnitude at each frequency where a series of individual spectra are added together and the sum is divided by the number of spectra.

5.3. Enveloping

The purpose of enveloping is to enhance small signals. The method first separates higher frequency bearing signals from low frequency machine vibrations by band pass filtering. The measurement problem at this point, is to detect small amplitudes. A defect signal in the time domain is very narrow, resulting in an energy component spread over a wide frequency range; consequently the harmonic amplitudes of the defect frequency are very nearly buried in noise.

5.4. Cepstrum

Cepstrum is the name given to a range of techniques all involving functions which can be considered as a “spectrum of a logarithmic spectrum”. The application of the power cepstrum to machine fault detection is based on the ability to detect the periodicity in the spectrum i.e. family of the uniformly spaced harmonics and side bands while being insensitive to the transmission path of the signal from an internal source to an external measurement point. The value of the main cepstrum peak was shown to be an excellent trend parameters; as it represents the average over a large number of individual harmonics, fluctuations in latter (for example as a result of load variations) were largely averaged out in the cepstrum value which gave a smooth trend curve with time.

References

[1] J. Ma, C.J. Li, Detection of localized defect in rolling element bearing via composite hypothesis test, Mechanical System and Signal Processing 9 (1995) 63-75.

[2] Y.W. Kim, G. Rizzoni, V.I. Utkin, Developing a fault tolerant power-train contrl system by integrating design of control and diagnostics, International Jurnal of Robust and Nonlinear Control 11 (2001) 1095-1114.

[3] H. Sohn, K. Wordwn, C.R. Farrar, Statistical damage classification under changing environmental and operation conditions, Journal of Intelligent Material System and Structures 13 (2002) 561-574.

[4] M. Nyberg, A General framework for fault diagnosis based on statistical hypothesis testing, in Twelfth International Workshop on Principles of Diagnosis (DX2001), Via Lattea, Italian Alps, 2001, pp. 135-142.

[5] M.L Fugate, H. Sohn C. R. Farrar, Vibration-based damage detection using statistical process control, Mechanical System and Signal Processing 15

(2001) 707-721.

[6] V.A. Skormin, L.J. Popyack, V.I. Gorodetski, M.L. Araiza, J.D. Michel, Application of cluster analysis in diagnostic related problems, in Proceeding of the 1999 IEEE Aerospace Conference, vol. 3, Snowmass at Aspen, CO, USA, 1999, pp. 161-168.

[7] M. artes, L. Del Castillo, J. Perez, Failure prevention and diagnosis in machine elements using cluster, in Proceeding of the Tenth International Congress on Sound and Vibration, Stockholm, Sweden, 2003, pp. 1197-1203.

[8] J. Schurmann, Pattern Recognition: A Unified View of Statistical and Neural Approaches, Wiley, New York, 1996.

[9] H. Ding, X. Gui, S. Yang, An approach to state recognition and knowledge based diagnosis for engines, Mechanical Systems and Signal Processing 5 (1991) 257–266.

[10] W.J. Staszewski, K. Worden, G.R. Tomlinson, Time–frequency analysis in gearbox fault detection using the Wigner–Ville distribution and pattern recognition, Mechanical Systems and Signal Processing 11 (1997) 673–692.

[11] S.K. Goumas, M.E. Zervakis, G.S. Stavrakakis, Classification of washing machines vibration signals using discrete wavelet analysis for feature extraction, IEEE Transactions on Instrumentation and Measurement 51 (2002) 497–508.

[12] X. Lou, K.A. Loparo, Bearing fault diagnosis based on wavelet transform and fuzzy inference, Mechanical Systems and Signal Processing 18 (2004) 1077 1095.

[13] M.-C. Pan, P. Sas, H. Van Brussel, Machine condition monitoring using signal classification techniques, Journal of Vibration and Control 9 (2003) 1103 1120.

[14] A.R. Webb, Statistical Pattern Recognition, Wiley, West Sussex, England,

2002.

[15] C.K. Mechefske, J. Mathew, Fault detection and diagnosis in low speed rolling element bearing. Part II: The use of nearest neighbour classification, Mechanical Systems and Signal Processing 6 (1992) 309–316.

[16] Q. Sun, P. Chen, D. Zhang, F. Xi, Pattern recognition for automatic machinery fault diagnosis, Journal of Vibration and Acoustics, Transactions of the ASME 126 (2004) 307–316.

[17] S. Poyhonen, P. Jover, H. Hyotyniemi, Signal processing of vibrations for condition monitoring of an induction motor, in: ISCCSP: 2004 First International Symposium on Control, Communications and Signal Processing, New York, 2004, pp. 499–502.

[18] M. Guo, L. Xie, S.-Q. Wang, J.-M. Zhang, Research on an integrated ICA SVM based framework for fault diagnosis, in: Proceedings of the 2003 IEEE International Conference on Systems, Man and Cybernetics, vol. 3, Washington, DC, USA, 2003, pp. 2710–2715.

[19] A. Siddique, G.S. Yadava, B. Singh, Applications of artificial intelligence techniques for induction machine stator fault diagnostics: Review, in: Proceedings of the IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives, New York, 2003, pp. 29–34.

[20] M.J. Roemer, C. Hong, S.H. Hesler, Machine health monitoring and life management using finite element-based neural networks, Journal of Engineering for Gas Turbines and Power—Transactions of the ASME 118 (1996) 830–835.

[21] E.C. Larson, D.P. Wipf, B.E. Parker, Gear and bearing diagnostics using neural network-based amplitude and phase demodulation, in: Proceedings of the 51st Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, VA, 1997, pp. 511–521.

[22] B. Li, M.-Y. Chow, Y. Tipsuwan, J.C. Hung, Neural-network-based motor rolling bearing fault diagnosis, IEEE Transactions on Industrial Electronics 47 (2000) 1060–1069.

[23] Y. Fan, C.J. Li, Diagnostic rule extraction from trained feedforward neural networks, Mechanical Systems and Signal Processing 16 (2002) 1073–1081.

[24] B.A. Paya, I.I. Esat, M.N.M. Badi, Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor, Mechanical Systems and Signal Processing 11 (1997) 751–765.

[25] B. Samanta, K.R. Al-Balushi, Artificial neural network based fault diagnostics of rolling element bearings using time-domain features, Mechanical Systems and Signal Processing 17 (2003) 317–328.

[26] J.K. Spoerre, Application of the cascade correlation algorithm (CCA) to bearing fault classification problems, Computers in Industry 32 (1997) 295–304.

[27] D.W. Dong, J.J. Hopfield, K.P. Unnikrishnan, Neural networks for engine fault diagnostics, in: Neural Networks for Signal Processing VII, New York, 1997, pp. 636–644.

[28] C.J. Li, T.-Y. Huang, Automatic structure and parameter training methods for modeling of mechanical systems by recurrent neural networks, Applied Mathematical Modelling 23 (1999) 933–944.

[29] P. Deuszkiewicz, S. Radkowski, On-line condition monitoring of a power transmission unit of a rail vehicle, Mechanical Systems and Signal Processing 17 (2003) 1321–1334.

[30] C.-C. Wang, G.-P.J. Too, Rotating machine fault detection based on HOS and artificial neural networks, Journal of Intelligent Manufacturing 13 (2002) 283–293.

[31] R.M. Tallam, T.G. Habetler, R.G. Harley, Self-commissioning training

algorithms for neural networks with applications to electric machine fault diagnostics, IEEE Transactions on Power Electronics 17 (2002) 1089–1095.

[32] Y.H. Yoon, E.S. Yoon, K.S. Chang, Process fault diagnostics using the integrated graph model, in: On-Line Fault Detection and Supervision in the Chemical Process Industries, Oxford, 1993, pp. 89–94.

[33] C.H. Hansen, R.K. Autar, J.M. Pickles, Expert systems for machine fault diagnosis, Acoustics Australia 22 (1994) 85–90.

[34] M.F. Baig, N. Sayeed, Model-based reasoning for fault diagnosis of twin spool turbofans, Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering 212 (1998) 109–116.

[35] Z.Y. Wen, J. Crossman, J. Cardillo, Y.L. Murphey, Case base reasoning in vehicle fault diagnostics, in: Proceedings of the International Joint Conference on Neural Networks 2003, Vols. 1–4, New York, 2003, pp. 2679–2684.

[36] M. Bengtsson, E. Olsson, P. Funk, M. Jackson, Technical design of condition based maintenance system—A case study using sound analysis and case-based reasoning, in: Maintenance and Reliability Conference—Proceedings of the Eighth Congress, Knoxville, USA, 2004.

[37] M.L. Araiza, R. Kent, R. Espinosa, Real-time, embedded diagnostics and prognostics in advanced artillery systems, in: 2002 IEEE Autotestcon Proceedings, Systems Readiness Technology Conference, New York, 2002, pp. 818–841.

[38] D.L. Hall, R.J. Hansen, D.C. Lang, The negative information problem in mechanical diagnostics, Transactions of the ASME Journal of Engineering for Gas Turbines and Power 119 (1997) 370–377.

[39] M. Stanek, M. Morari, K. Frohlich, Model-aided diagnosis: An inexpensive combination of model-based and case-based condition assessment, IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and

Reviews 31 (2001) 137–145.

[40] L. B. Jack, A. K. Nandi, Fault detection using support vector machine and artificial neural network, augmented by genetic algorithm, Mechnaical System and Signal Processing 16 (2002) 373-390.

[41] W. Yan, H. Shao, Application of support vector machine nonlinear classifier to fault diagnosis, Proceeding of the World Congress on Intelligent Control and Automation, 2002, pp. 2697-2700.

[42] B. Samanta, K.R. Al-Balushi, S.A Al-Araimi, Artificial neural network and support vector machine with genetic algorithm for bearing fault detection, Engineering Application of Artificial Intelligence 16 (2003) 657-665.

[43] B. Samanta, Gear fault detection using artificial neural networks and support vector machines with genetic algorithms, Mechanical Systems and Signal Processing 18(3) (2004) 625-644.

[44] A. Rojas, A.K. Nandi, Practical scheme for fast detection and classification of rolling-element bearing faults using support vector machines, Mechanical Systems and Signal Processing 20(7) (2006) 1523-1536.

[45] A. Rojas, K. Nandi, Detection and classification of rolling-element bearing faults using support vector machines, IEEE Workshop on Machine Learning for Signal Processing, 2005, pp.153-158.

[46] Z. Zhang, Wenzhi, M. Shen, Active learning of support vector machine for fault diagnosis of bearings, Lecture Notes in Computer Science 3973 (2006) 390-395.

[47] V. Sugumuran, V. Muralidharan, K.I. Ramachandran, Feature selection using decision tree and classification proximal support vector machine for fault diagnostic of roller bearing, Mechanical System and Signal Processing 21(2) (2007) 930-942.

[48] Q. Hu, Z. He, Z. Zhang, Y. Zi, Fault diagnosis of rotating machinery based on improved wavelet package transform and SVM ensemble, Mechanical

System and Signal Processing 21(2) (2007) 688-705.

[49] S. Pöyhönen, M. Negrea, A. Arkkio, H. Hyotyniemi, H. Koivo, Fault diagnosis of an electrical machine with multiple support vector classifiers, Proceeding of the IEEE International Symposium on Intelligent Control, 373-378, 2002.

[50] S. Pöyhönen, M. Negrea, A. Arkkio, H. Hyotyniemi, H. Koivo, Coupling pairwise support vector machines for fault classification, Control Engineering Practice 13(6) (2005) 759-769.

[51] C. Zhitong, F. Jiazhong, C. Hongpingn, H. Guoguang, E. Ritchie, Support vector machine used to diagnose the fault of rotor broken bars of induction motors, 6th International Conference on Electrical Machines and System (ICEMS), 891-894, 2003.

[52] R. Fang, Induction machine rotor diagnosis using support vector machines and rough set, LNAI 4114, 631-636, 2006.

[53] A. Widodo, B.S. Yang, T. Han, Combination of independent component analysis and support vector machine for intelligent faults diagnosis of induction motors, Expert System with Application 32 (2007) 299-312.

[54] A. Widodo, B.S. Yang, T. Han, D.J. Kim, Fault diagnosis of induction motor using independent component analysis and multi-class support vector machine, Proceeding of the 11th Asia-Pacific Vibration Conference, 144-149, 2005.

[55] A. Widodo, B.S. Yang, Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors, Expert System with Application 33(1) (2007) 241-250.

[56] A. Widodo, B.S. Yang, Faults detection and classification of induction motor using wavelet support vector machine, Korean Society of Power System Engineering (KSPSE) Conference, 79-84, 2006.

[57] A. Widodo, B.S. Yang, Wavelet support vector machine for machine faults

classification, Lecture Series on Computer and Computational Sciences 8 (2007) 1-5.

[58] R. Ramesh, M.A. Mannan, A.N. Poo, S.S. Keerthi, Thermal error measurement and modeling in machine tools, Part II. Hybrid Bayesian network-support vector machine model, International Journal of Machine Tools & Manufacture 43 (2003) 405-419.

[59] J. Sun, M. Rahman, Y.S. Wong, G.S. Hong, Multiclassification of tool wear with support vector machine by manufacturing loss consideration, International Journal Machine Tools & Manufacture 44 (2004) 1179-1187.

[60] J. Sun, G.S. Hong, M. Rahman, Y.S. Wong, The application of nonstandard support vector machine in tool condition monitoring system, Proceeding of the 2nd IEEE International Workshop on Electronic Design, Test and Applications, 1-6, 2004.

[61] S. Cho, S. Asfour, A. Onar, N. Kaundinya, Tool breakage detection using support vector machine learning in a milling process, International Journal of Machine Tools & Manufacture 45 (2005) 241-249.

[62] D.M.J. Tax, A. Ypma, R.P.W. Duin, Pump failure determination using support vector data description, Lecture Notes in Computer Science, 415-425, 1999.

[63] J. Gao, W. Shi, J. Tan, F. Zhong, Support vector machine based approach for fault diagnosis of valves in reciprocating pumps, Proceeding of the IEEE Canadian Conference on Electrical & Computer Engineering, 1622-1627, 2002.

[64] S.F. Yuan, F.L. Chu, Support vector machines-based fault diagnosis for turbo-pump rotor, Mechanical System and Signal Processing 20 (4) (2006) 939-952.

[65] S. F Yuan, F.L Chu, Fault diagnosis based on support vector machine with parameter optimization by artificial immunization algorithm, Mechanical

System and Signal Processing 21(3) (2007) 1318-1330.

[66] B.S. Yang, W.W. Hwang, D.J. Kim, A.C.C. Tan, Condition classification of small reciprocating compressor for refrigerators using artificial neural networks and support vector machines, Mechanical System and Signal Processing 19 (2) (2005) 371-390.

[67] B.S. Yang, W.W. Hwang, M.H. Ko, S.J. Lee, Cavitation detection of butterfly valve using support vector machines, Journal of Sound Vibration 287(1-2) (2005) 25-43.

[68] Q. Ren, X. Ma, G. Miao, Application of support vector machine in reciprocating compressor valve fault diagnosis, Lecture Notes in Computer Science (2005) 81-84.

[69] Y. Li, Y.Z. Chai, R.P. Yin, X.M. Xu, Fault diagnosis based on support vector machine ensemble, Proceeding of the 4th International Conference on Machine Learning and Cybernetics, 2005, pp. 3309-3314.

[70] Z. Zhang, Q. Hu, Z. He, Fuzzy support vector machine and its application to mechanical condition monitoring, Lecture Notes in Computer Science 3496 (2005) 937-942.

[71] C. Batur, C.C. Chan, Support vector machine for fault detection, Proceeding of the 41th IEEE Conference on decision and Control, 2002, pp. 1355-1356.

[72] L. Han, L Ding, J. Yu, Q. Li, Y. Liang, Power plant boiler air preheater hot spots detection system based on least square support vector machine, Lecture Notes in Computer Science (2004) 598-604,.

[73] K.H. Choi, S.M. namburu, M.S. Azam, J. Luo, K.R. Pattipati, A.P. Hine, Fault diagnosis in HVAC chillers, IEEE Instrumentation & Measurement Magazine, 24-32, Aug. 2005.

[74] M. Rychetsky, S. Ortmann, M. Glesner, Support vector for engine knock detection, Proceeding of the International Joint Conference on Neural Network, 1999, pp. 969-974.

[75] Y. Xu, L. Wang, Fault diagnosis system based on rough set theory and support vector machine, LNAI 3614 (2005) 980-988.

[76] Z. Hu, Y. Cai, X. He, X. Xu, Fusion of multi-class support vector machine for fault diagnosis, Proceeding of the American Control Conference, 2005, pp. 1941-1945.

[77] M. Ge, R. Du, G. Zhang, Y. Xu, Fault diagnosis using support vector machine with an application in sheet metal stamping operations, Mechanical System and Signal Processing 18(1) (2004) 143-159.

[78] T. Han, Development of a feature-based fault diagnostics system and its application to induction motors, PhD Thesis, Department of Mechanical Engineering, Pukyong National University, South Korea.

[79] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, Oxford, 1995.

[80] R.O. Duda, P.E. Hart, D.G. Stock, Pattern Classification, John Wiley & Sons, New York, 2001.

[81] B.S. Yang, T. Han, W.W. Hwang, Fault diagnosis of rotating machinery based on multi-class support vector machines, Journal of Mechanical Science and Technology 19(3) (2005) 845-858.

[82] B.S. Yang, D.S Lim, J.L. An, Vibration diagnostic system of rotating machinery using artificial neural network and wavelet transfom, Proceeding of 13th International Congress on COMADEM, Houston, USA, 2000. pp. 12-20.

[83] B.S. Yang, T. Han, J.L. An, ART-KOHONEN neural network for faults diagnosis of rotating machinery, Mechanical System and Signal Processing 18(3) (2004) 645-657.

[84] B.S. Yang, S.K. Jeong, Y.M. Oh, A.C.C. Tan, Case-based reasoning system with Petri nets for induction motor fault diagnosis, Expert Systems with Applications 27(2) (2004) 301-311.

III. Component Analysis and Support Vector Machine

1. Introduction

Component analysis is a technique of multivariate statistical analysis that can linearly or nonlinearly transforms an original set variables into a substantially smaller set variables. It can be viewed as a classical method of multivariate statistical analysis for dimensionality reduction. Because of the fact that a small set of uncorrelated or independent variables is much easier to understand and use in further analysis than a larger set of correlated or dependent variables. This technique has been widely applied to virtually every substantive area including cluster analysis, visualization of high-dimensionality data, regression, data compression and pattern recognition. In this research, component analysis is used to extract the sensitive feature from original features.

Support vector machine (SVM) is a kind of machine learning based on statistical learning theory which can be applied to pattern classification. SVM becomes famous and popular in machine learning community due to the excellence of generalization ability than the traditional method such as neural network. Therefore, SVM has been successfully applied to a number of applications ranging from face detection, verification, and recognition, object detection and recognition, handwritten character and digit recognition, text detection and categorization, speech and speaker verification, recognition, information and image retrieval, prediction and so on.

2. Dimensionality Reduction using Component Analysis

Dimensionality reduction is one of the important preprocessing steps in high-dimensional data analysis. The goal of dimensionality reduction is to embed high-dimensional data samples in a low-dimensional space while most of intrinsic information contained in the data is preserved. Once dimensionality reduction is carried out appropriately, we can utilize the compact representation of the data for various succeeding tasks such as visualization, classification, etc.

Usually, somebody who works in pattern recognition area will face the high dimensionality of data, namely curse of dimensionality. It means that the processing of the data will be slow and requires a lot of memory and time. The other problem with high dimensionality of data is the classification using some algorithms will overfit to the data training. Thus, it leads to poor generalization to the training samples. Feature extraction is a general term for methods for constructing combinations of the variables which get around above problems but still describe the data sufficiently accurately. Here, several methods of feature extraction technique will be discussed to give the understanding of its process.

Component analysis is an unsupervised approach to finding the good features from the data. In unsupervised learning or clustering there is no explicit teacher, and the system forms clusters or natural grouping of the input patterns. In this section, component analysis using linear and nonlinear technique will be introduced. Component analysis has objective to project the high-dimensional data onto a lower dimensional space. Thus, component analysis has capability to reduce the dimensionality by combining the features.

2.1. Linear Technique

2.1.1. Principal Component Analysis (PCA)

Principal component analysis (PCA) has been called one of the most valuable results from applied linear algebra. PCA is used abundantly in all forms of analysis-from neuroscience to computer graphics-because it is a simple, non-parametric method of extracting relevant information from confusing data sets. With minimal additional effort PCA provides a roadmap for how to reduce a complex data set to a lower dimension to reveal the sometimes hidden, simplified structure that often underlie it.

Moreover, PCA is a useful statistical technique that has found in many fields, such as face recognition, optical character and speech recognition [1-3]. It is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since finding patterns in data are difficult in high dimensional condition, where the luxury of graphical representation is not available, PCA is a convenience tool for analyzing data. The other main advantage of PCA is that there is no much loss information when the data are compressed. Principal components (PC) are uncorrelated and ordered such that the *k*th PC has the *k*th largest variance among all PC. The *k*th of PC can be interpreted as the direction that maximizes the variation of the projection of the data points such that it is orthogonal to the first *k-1* PC.

Given p dimensionality data set $\mathbf{x}_i$, the m principal axis T_1, T_2,…, T_m where $1 \le m \le p$ are orthogonal axis onto which the retained variance is maximum in the projected space. Generally, T_1, T_2,…, T_m can be given by the m leading eigenvectors of covariance matrix

$$\mathbf{S} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^T (x_i - \mu) \tag{3.1}$$

where $x_i \in \mathbf{x}_i$, N is the number of samples, so that

$$\mathbf{S}T_i = \lambda_i T_i \qquad i = 1,\ldots,m \tag{3.2}$$

where λ_i is the ith largest eigenvalue of **S**. The m principal components of a given observation vector $\mathbf{x} \in \mathbf{x}_i$ are given by

$$\mathbf{y} = [y_1, ..., y_m] = [T_1^T x, ..., T_m^T x] = \mathbf{T}^T \mathbf{x} \tag{3.3}$$

The m principal components of the **x** are the uncorrelated in the projected space. In multi-class problem, the variations of data are determined on a global basis that is the principal axis are derived from a global covariance matrix

$$\hat{\mathbf{S}} = \frac{1}{N} \sum_{j=1}^{K} \sum_{i=1}^{N} (x_i - \hat{\mu})^T (x_i - \hat{\mu}) \tag{3.4}$$

where $\hat{\mu}$ is the global mean of all samples, K is the number of class, N_j is the number of samples in class j. The principle axis T_1, T_2,…, T_m are therefore the m leading eigenvectors of $\hat{\mathbf{S}}$

$$\hat{\mathbf{S}} T_i = \hat{\lambda}_i T_i \qquad i = 1, ..., m \tag{3.5}$$

where $\hat{\lambda}_i$ is the ith eigenvalue of $\hat{\mathbf{S}}$.

An assumption made for dimensionality reduction by PCA is that most information of the observation vectors is contained in the subspace spanned by the first m principal axis. Therefore, each original data vector can be represented by its principal component vector

$$\mathbf{y} = \mathbf{T}^T \mathbf{x} \tag{3.6}$$

where $\mathbf{T} = [T_1, T_2, \ldots, T_m]$.

The principal components of PCA are uncorrelated and they have sequentially maximum variances. The significant property is that the mean squared approximation error in the representation of the original inputs by the first several principal components is minimal [4].

2.1.2. 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. Recently this technique has been demonstrated to be able to extract independent components from the mixed signals. Here independence means the information carried by one component can not be inferred by the others. Statistically this means that joint probability of independent quantities is obtained as the product of the probability of each of them. A generic ICA model can be written as

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{3.7}$$

where **A** is an unknown full-rank matrix, called the mixing matrix, and **s** is the independent component (IC) data matrix, and **x** is the measured variable data matrix. The basic problem of ICA is to estimate the independent component matrix **s** or to estimate the mixing matrix **A** from the measured data matrix **x** without any knowledge of **s** or **A**.

The ICA algorithm normally finds the independent components of a data set by minimizing or maximizing some measure of independence. Cardoso [5] gave a review of the solution to the ICA problem using various information theoretic criteria, such as mutual information, negentropy, and maximum entropy, as well as maximum likelihood approach. The fixed-point algorithm used due to its suitability for handling raw time domain data and good convergence properties. This algorithm will now be described briefly.

The first step is to pre-whiten the measured data vector **x** by a linear transformation, to produce a vector $\tilde{\mathbf{x}}$ whose elements are mutually uncorrelated and all have unit variance. Singular value decomposition (SVD) of the covariance matrix $\mathbf{C} = E[\mathbf{x}\mathbf{x}^T]$ yields

$$\mathbf{C} = \mathbf{\Psi\Sigma\Psi}^T \tag{3.8}$$

where $\sum = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$ is a diagonal matrix of singular values and Ψ is the associated singular vector matrix. Then, the vector $\tilde{\mathbf{x}}$ can be expressed as

$$\tilde{\mathbf{x}} = \mathbf{\Psi\Sigma}^{-1/2}\mathbf{\Psi}^T\mathbf{x} = \mathbf{QAs} = \mathbf{Bs} \tag{3.9}$$

where **B** is an orthogonal matrix as verified by the following relation:

$$E[\tilde{\mathbf{x}} \cdot \tilde{\mathbf{x}}T] = \mathbf{B}E[\mathbf{s} \cdot \mathbf{s}^T]\mathbf{B}^T = \mathbf{B}\mathbf{B}^T = \mathbf{I} \tag{3.10}$$

An advantage of using an SVD-based technique is the possibility of noise reduction by discarding singular values smaller than a given threshold. We have therefore reduced the problem of finding an arbitrary full-rank matrix **A** to the simpler problem of finding an orthogonal matrix **B** since **B** has fewer parameters to estimate as a result of the orthogonality constraint.

The second step is to employ the fixed point algorithm. Define a separating matrix **W** that transform the measured data vector **x** to a vector **y**, such that all elements $\mathbf{y}_i$ are both mutually correlated and have unit variance. The fixed-point algorithm then determines **W** by maximizing the absolute value of kurtosis of **y**. The vector **y** has the properties required for the independent components, thus

$$\tilde{\mathbf{s}} = \mathbf{y} = \mathbf{W}\mathbf{x} \tag{3.11}$$

From Eq. (3.9), we can estimate **s** as follows

$$\tilde{\mathbf{s}} = \mathbf{B}^T \tilde{\mathbf{x}} = \mathbf{B}^T \mathbf{Q}\mathbf{x} \tag{3.12}$$

From Eqs. (3.11) and (3.12) the relation of **W** and **B** can be expressed as

$$\mathbf{W} = \mathbf{B}^T \mathbf{Q} \tag{3.13}$$

To calculate **B**, each column vector $\mathbf{b}_i$ is initialized and then updated so that *i*th independent component $\mathbf{s}_i = (\mathbf{b}_i)^T \tilde{\mathbf{x}}$ may have great non-Gaussianity. Hyvärinen and Oja [6] showed that non-Gaussian represents independence using the central limit theorem. There are two common measures of non-Gaussianity: kurtosis and negentropy. Kurtosis is sensitive to outliers. On the other hand, negentropy is based on the information theoretic quantity of (differential) entropy. Entropy is a measure of the average uncertainty in a random variable and the differential entropy *H* of random variable *y* with density *f*(*y*) is defined as

$$H(y) = -\int f(y) \log f(y) dy \tag{3.14}$$

A Gaussian variable has maximum entropy among all random variables with equal variance [6]. In order to obtain a measure of non-Gaussianity that is zero for a Gaussian variable, the negentropy J is defined as follows

$$J(y) = H(y_{gauss}) - H(y) \tag{3.15}$$

where y_{gauss} is a Gaussian random variable with the same variance as y. Negentropy is nonnegative and measures the departure of y from Gaussianity. However, estimating negentropy using Eq. (3.15) would require an estimate of the probability density function. To estimate negentropy efficiently, simpler approximations of negentropy suggested as follows:

$$J(y) \approx [E\{G(y)\} - E\{(\upsilon)\}]^2 \tag{3.16}$$

where y is assumed to be of zero mean and unit variance, v is a Gaussian variable of zero mean and unit variance, and G is any non-quadratic function. By choosing G wisely, one obtains good approximations of negentropy. A number of functions for G are:

$$G_1(\upsilon) = \frac{1}{a_1}\log\cosh(a_1\upsilon) \tag{3.17}$$

$$G_2(\upsilon) = \exp(-a_2\upsilon^2/2) \tag{3.18}$$

$$G_3(\upsilon) = \upsilon^4 \tag{3.19}$$

where $1 \leq a_1 \leq 2$ and $a_2 \approx 1$. Among these three functions, G_1 is a good general-purpose contrast function and was therefore selected for use in the present study.

Based on approximate form for the negentropy, Hyvärinen [7] introduced a very simple and highly efficient fixed-point algorithm for ICA, calculated over sphered zero-mean vector $\tilde{\mathbf{x}}$. This algorithm calculates one column of the matrix **B** and allows the identification of one independent component; the corresponding independent component can then be found using Eq. (3.12). The algorithm is repeated to calculate each independent component.

2.2. Nonlinear Technique

2.2.1. Kernel PCA

Kernel PCA is one approach of generalizing linear PCA into nonlinear case using the kernel method. The idea of kernel PCA is to firstly map the original input vectors $\mathbf{x}_i$ into a high-dimensional feature space $\phi(\mathbf{x}_i)$ and than calculate the linear PCA in $\phi(\mathbf{x}_i)$ [8]. By mapping $\mathbf{x}_i$ into $\phi(\mathbf{x}_i)$ whose dimension is assumed to larger than the number of training samples m, kernel PCA solves the eigenvalue problem of Eq. (3.2)

$$\hat{\mathbf{S}}\mathbf{T}_i = \hat{\lambda}_i \mathbf{T}_i \qquad i = 1,\ldots,m \tag{3.20}$$

where $\hat{\mathbf{S}}$ is the sample covariance matrix of $\phi(\mathbf{x}_i)$, $\hat{\lambda}_i$ is one of the non-zero eigenvalues of $\hat{\mathbf{S}}$ and $\mathbf{T}_i$ is the corresponding eigenvectors. The $\hat{\mathbf{S}}$ on the feature space can be constructed by

$$\hat{\mathbf{S}} = \frac{1}{m}\sum_{i=1}^{m}\phi(\mathbf{x}_i)\phi(\mathbf{x}_i)^T \tag{3.21}$$

From Eq. (3.21), we can obtain the non-zero eigenvalues that are positive. Let us define matrix $\mathbf{Q}$ as

$$\mathbf{Q} = [\phi(\mathbf{x}_1),\ldots,\phi(\mathbf{x}_m)] \tag{3.22}$$

Then Eq. (3.21) can be expressed by

$$\hat{\mathbf{S}} = \frac{1}{m}\mathbf{Q}\mathbf{Q}^T \tag{3.23}$$

Moreover, we can construct a Gram matrix using Eq. (3.22) which is their element can be determined by kernel

$$\mathbf{R} = \mathbf{Q}^T\mathbf{Q} \tag{3.24}$$

$$\mathbf{R}_{ij} = \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_i) = (\phi(\mathbf{x}_i)\cdot\phi(\mathbf{x}_j)) \quad = K(\mathbf{x}_i,\mathbf{x}_j) \tag{3.25}$$

Denote $\mathbf{V} = (\gamma_1, \gamma_2, \ldots, \gamma_m)$ and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$ are eigenvectors and eigenvalues of $\mathbf{R}$ respectively, we can calculate the orthonormal eigenvectors $\boldsymbol{\beta}_j$ as

$$\boldsymbol{\beta}_j = \frac{1}{\sqrt{\lambda_m}}\mathbf{Q}\boldsymbol{\gamma} \qquad j = 1,...,l \tag{3.26}$$

Then we define matrix **B** as

$$\mathbf{B} = (\beta_1, \beta_2, ..., \beta_l) = \mathbf{Q}\mathbf{V}\boldsymbol{\Lambda}^{-1/2} \tag{3.27}$$

The whitening matrix **P** can be derived from Eq. (3.27) and expressed by

$$\mathbf{P} = \mathbf{B}\left(\frac{1}{m}\boldsymbol{\Lambda}\right)^{-1/2} = \sqrt{m}\mathbf{Q}\mathbf{V}\boldsymbol{\Lambda}^{-1} \tag{3.28}$$

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

$$\begin{aligned}\mathbf{r} = \mathbf{P}^T\phi(\mathbf{x}_i) &= \sqrt{m}\boldsymbol{\Lambda}^{-1}\mathbf{V}^T\mathbf{Q}^T\phi(\mathbf{x}) = \sqrt{m}\boldsymbol{\Lambda}^{-1}\mathbf{V}^T[K(\mathbf{x}_1,\mathbf{x}), K(\mathbf{x}_2,\mathbf{x}),...,K(\mathbf{x}_m,\mathbf{x})) \\ &= \sqrt{m}\boldsymbol{\Lambda}^{-1}\mathbf{V}^T\mathbf{R}\end{aligned} \tag{3.29}$$

2.2.2. Kernel ICA

Practically speaking, the kernel ICA is the combination of centering and whitening process using kernel PCA as previously explanation and iterative section using ICA. The following task is to find the mixing matrix **W** in the kernel PCA-transformed space to recover independent components $\tilde{\mathbf{s}}$ from **r,** recall Eq. (3.11)

$$\tilde{\mathbf{s}} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{r} \tag{3.30}$$

There are many algorithms to perform ICA. In this study, we employ the second order of ICA, proposed by Belouchrani et al. [9] which is adopted in ICALAB toolbox [10]. In summary, the nonlinear feature extraction using kernel ICA in this dissertation performs two phases: whitened process using kernel PCA and ICA transformation in the kernel PCA whitened space.

3. Support Vector Machine (SVM)

3.1. Overview

Support vector machine (SVM) is a relatively new computational learning method based on the statistical learning theory. Introduced by Vapnik and his co-workers [11-13], SVM becomes famous and popular in machine learning community due to the excellence of generalization ability than the traditional method such as neural network. Therefore, SVM have been successfully applied to a number of applications ranging from face detection, verification, and recognition, object detection and recognition, handwritten character and digit recognition, text detection and categorization, speech and speaker verification, recognition, information and image retrieval, prediction and so on.

In machine condition monitoring and fault diagnosis problem, SVM is employed for recognizing special patterns from acquired signal, and then these patterns are classified according to the fault occurrence in the machine. After signal acquisition, a feature representation method can be performed to define the features e.g. statistical feature of signal for classification purposes. These features can be considered as patterns that should be recognized using SVM.

3.2. Basic Theory: Binary Classification Using SVM

Given data input $\mathbf{x}_i$ (i = 1, 2, …, M), M is the number of samples. The samples are assumed have two classes namely positive class and negative class. Each of classes associate with labels be $y_i = 1$ for positive class and $y_i = -1$ for negative class, respectively. In the case of linearly data, it is possible to determine the hyperplane $f(\mathbf{x}) = 0$ that separates the given data

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = \sum_{j=1}^{M} w_j x_j + b = 0 \tag{3.31}$$

where **w** is M–dimensional vector and b is a scalar. The vector **w** and scalar b are used to define the position of separating hyperplane. The decision function is made using sign $f(x)$ to create separating hyperpline that classify input data in either positive class and negative class.

A distinctly separating hyperplane should be satisfy the constraints

$$\begin{aligned} f(x_i) &= 1 \quad \text{if } y_i = 1 \\ f(x_i) &= -1 \quad \text{if } y_i = -1 \end{aligned} \tag{3.32}$$

or it can be presented in complete equation

$$y_i f(\mathbf{x}_i) = y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 \quad \text{for} \quad i = 1,2,\ldots,M \tag{3.33}$$

The separating hyperplane that creates the maximum distance between the plane and the nearest data, i.e., the maximum margin, is called the optimal separating hyperplane. An example of the optimal hyperplane of two data sets is presented in Fig. 3.1.

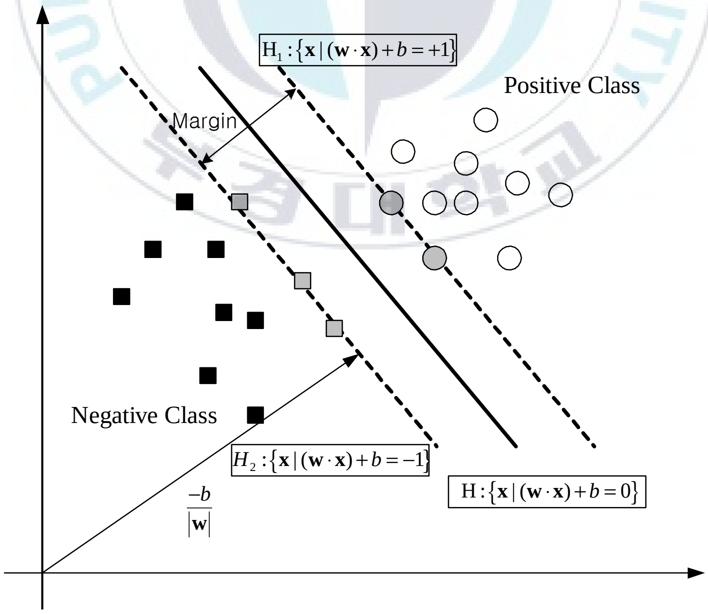


Fig. 3.1 Classification of two classes using SVM.

In Fig. 3.1, the series data points for two different classes of data are shown, black squares for negative class and white circles for positive class. The SVM try to place a linear boundary between the two different classes, and orientate it in such way that the margin represented by the dotted line is maximized. Furthermore, SVM attempts to orientate the boundary to ensure that the distance between the boundary and the nearest data point in each class is maximal. Then, the boundary is placed in the middle of this margin between two points. The nearest data points that used to define the margin are called support vectors, represented by the grey circles and squares. When the support vectors have been selected the rest of the feature set is not required, as the support vectors can contain all the information based need to define the classifier. From the geometry the geometrical margin is found to be $\|\mathbf{w}\|^{-2}$.

Taking into account the noise with slack variables ξ_i and the error penalty C, the optimal hyperplane separating the data can be obtained as a solution to the following optimization problem

$$\text{minimize} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{M}\xi_i \tag{3.34}$$

$$\text{subject to} \quad \begin{cases} y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i, & i = 1,\dots,M \\ \xi_i \geq 0 & i = 1,\dots,M \end{cases} \tag{3.35}$$

where ξ_i is measuring the distance between the margin and the examples $\mathbf{x}_i$ that lying on the wrong side of the margin. The calculation can be simplified by converting the problem with Kuhn-Tucker condition into the equivalent Lagrangian dual problem, which will be

$$\text{minimize} \quad L(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{M}\alpha_i y_i(\mathbf{w}\mathbf{x}_i + b) + \sum_{i=1}^{M}\alpha_i \tag{3.36}$$

The task is minimizing Eq. (3.36) with respect to $\mathbf{w}$ and b, while requiring the derivatives of L to α to vanish. At optimal point, we have the following saddle point equations

$$\frac{\partial L}{\partial \mathbf{w}} = 0, \quad \frac{\partial L}{\partial b} = 0 \tag{3.37}$$

Which replace into form

$$\mathbf{w} = \sum_{i=1}^{M} \alpha_i y_i \mathbf{x}_i, \qquad \sum_{i=1}^{M} \alpha_i y_i = 0 \tag{3.38}$$

From Eq. (3.38), we find that **w** is contained in the subspace spanned by the $\mathbf{x}_i$. Using substitution Eq. (3.38) into Eq. (3.37), we get the dual quadratic optimization problem

$$\text{maximize} \quad L(\alpha) = \sum_{i=1}^{M} \alpha_i - \frac{1}{2} \sum_{i,j=0}^{M} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \mathbf{x}_j \tag{3.39}$$

subject to $\alpha_i \geq 0, \quad i = 1, \ldots, M.$

$$\sum_{i=1}^{M} \alpha_i y_i = 0 \tag{3.40}$$

Thus, by solving the dual optimization problem, one obtains the coefficients α_i which is required to express the **w** to solve Eq. (3.34). This leads to non-linear decision function.

$$f(\mathbf{x}) = sign\left(\sum_{i,j=1}^{M} \alpha_i y_i (\mathbf{x}_i \mathbf{x}_j) + b \right) \tag{3.41}$$

SVM can also be used in non-linear classification tasks with application of kernel functions. The data to be classified is mapped onto a high-dimensional feature space, where the linear classification is possible. Using the non-linear vector function $\mathbf{\Phi}(\mathbf{x}) = (\phi_1(\mathbf{x}), ..., \phi_l(\mathbf{x}))$ to map the n-dimensional input vector **x** onto l-dimensional feature space, the linear decision function in dual form is given by

$$f(\mathbf{x}) = sign\left(\sum_{i,j=1}^{M} \alpha_i y_i (\mathbf{\Phi}^T(\mathbf{x}_i)\mathbf{\Phi}(\mathbf{x}_j)) + b \right) \tag{3.42}$$

Working in the high-dimensional feature space enables the expression of complex functions, but it also generates the problem. Computational problem occur due to the large vectors and the overfitting also exists due to the high-dimensionality. The latter problem can be solved by using the kernel function. Kernel is a function that returns a dot product of the feature space mappings of the original data points, stated as $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{\Phi}^T(\mathbf{x}_i)\mathbf{\Phi}_j(\mathbf{x}_j))$. When applying a kernel function, the learning in the feature space does not require explicit evaluation of $\mathbf{\Phi}$ and the decision function will be

$$f(\mathbf{x}) = sign\left(\sum_{i,j=1}^{M} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_j) + b\right) \tag{3.43}$$

Any function that satisfies Mercer's theorem [11, 14] can be used as a kernel function to compute a dot product in feature space. There are different kernel functions used in SVM, such as linear, polynomial and Gaussian RBF. The selection of the appropriate kernel function is very important, since the kernel defines the feature space in which the training set examples will be classified. The definition of legitimate kernel function is given by Mercer's theorem. The function must be continuous and positive definite. In this work, linear, polynomial and Gaussian RBF functions were evaluated and formulated in Table 3.1.

Table 3.1 Formulation of kernel functions

Kernel	$K(\mathbf{x}, \mathbf{x}_j)$
Linear	$\mathbf{x}^{\mathrm{T}} \cdot \mathbf{x}_j$
Polynomial	$(\gamma\, \mathbf{x}^{\mathrm{T}} \cdot \mathbf{x}_j + \mathrm{r})^d \quad , \quad \gamma \; > 0$
Gaussian RBF	$\exp(-\, \|\mathbf{x} - \mathbf{x}_j\|^2 / 2\gamma^2)$

3.3. SVM Solver

3.3.1. Quadratic Programming (QP)

Vapnik [15] described a method which used the projected conjugate gradient algorithm to solve the SVM-QP problem, which has been known as *chunking*. The chunking algorithm uses the fact that the value of the quadratic form is the same if you remove the rows and columns of the matrix that corresponds to zero Lagrange multipliers. Therefore, chunking seriously reduces the size of the matrix from the number of training examples squared to approximately the number of non-zero Lagrange multipliers squared. However, chunking still cannot handle large-scale training problems, since even this reduced matrix cannot fit into memory. Osuna, Freund and Girosi [16] proved a theorem which suggests a whole new set of QP algorithms for SVM. The theorem proves that the large QP problem can be broken down into a series of smaller QP sub-problems.

3.3.2. Sequential Minimum Optimization (SMO)

Sequential minimal optimization (SMO) proposed by Platt [17] is a simple algorithm that can be used to solve the SVM-QP problem without any additional matrix storage and without using the numerical QP optimization steps. This method decomposes the overall QP problem into QP sub-problems using the Osuna's theorem to ensure convergence. In this dissertation, SMO is used as a solver and detail descriptions can be found in Platt [17].

In order to solve the two Lagrange multipliers α_1, α_2, SMO first computes the constraints on these multipliers and then solves for the constrained minimum. For convenience, all quantities that refer to the first multiplier will have a subscript 1, while all quantities that refer to the second multiplier will have a subscript 2. The new values of these multipliers must lie on a line in (α_1, α_2) space, and in the box defined by $0 \le \alpha_1, \alpha_2 \le C$.

$$\alpha_1 y_1 + \alpha_2 y_2 = \alpha_1^{old} y_1 + \alpha_2^{old} y_2 = \text{constant} \tag{3.44}$$

Without loss of generality, the algorithm first computes the second Lagrange multipliers α_2^{new} and successively uses it to obtain α_1^{new}. The box constraint $0 \le \alpha_1, \alpha_2 \le C$, together with the linear equality constraint $\Sigma\alpha_i y_i = 0$, provides a more restrictive constraint on the feasible values for α_2^{new}. The boundary of feasible region for α_2 can be applied as follows

$$\text{If } y_1 \neq y_2;\ L = \max(0,\ \alpha_2^{\text{old}} - \alpha_1^{\text{old}}),\ H = \min(C,\ C + \alpha_2^{\text{old}} - \alpha_1^{\text{old}}), \tag{3.45}$$

$$\text{If } y_1 = y_2;\ L = \max(0,\ \alpha_1^{\text{old}} + \alpha_2^{\text{old}} - C),\ H = \min(C,\ C + \alpha_1^{\text{old}} + \alpha_2^{\text{old}}) \tag{3.46}$$

The second derivative of the objective function along the diagonal line can be expressed as:

$$\eta = K(\mathbf{x}_1, \mathbf{x}_1) + K(\mathbf{x}_2, \mathbf{x}_2) - 2K(\mathbf{x}_1, \mathbf{x}_2). \tag{3.47}$$

Under normal circumstances, the objective function will be positive definite, there will be a minimum along the direction of the linear equality constraint, and η will be greater than zero. In this case, SMO computes the minimum along the direction of the constraint:

$$\alpha_2^{\text{new}} = \alpha_2^{\text{old}} + \frac{y_2\left(E_1^{\text{old}} - E_2^{\text{old}}\right)}{\eta} \tag{3.48}$$

where E_i is the prediction error on the ith training example. As a next step, the constrained minimum is found by clipping the unconstrained minimum to the ends of the line segment:

$$\alpha_2^{\text{new,clipped}} = \begin{cases} H & \text{if } \alpha_2^{\text{new}} \ge H; \\ \alpha_2^{\text{new}} & \text{if } L < \alpha_2^{\text{new}} < H; \\ L & \text{if } \alpha_2^{\text{new}} \le L; \end{cases} \tag{3.49}$$

Now, let $s = y_1 y_2$. The value of α_1^{new} is computed from the new α_2^{new}:

$$\alpha_1^{\text{new}} = \alpha_1^{\text{old}} + s(\alpha_2^{\text{old}} - \alpha_2^{\text{new}}) \tag{3.50}$$

Solving Eq. (3.39) for the Lagrange multipliers does not determine the threshold b of the SVM, so b must be computed separately. The following threshold b_1, b_2 are valid when the new α_1, α_2 are not at the each bounds, because it forces the output of the SVM to be y_1, y_2 when the input is $\mathbf{x}_1$, $\mathbf{x}_2$ respectively

$$b_1 = E_1 + y_1\,(\alpha_1^{\text{new}} - \alpha_1^{\text{old}})\,K(\mathbf{x}_1, \mathbf{x}_1) + y_2\,(\alpha_2^{\text{new,clipped}} - \alpha_2^{\text{old}})\,K(\mathbf{x}_1, \mathbf{x}_2) + b^{\text{old}} \tag{3.51}$$

$$b_2 = E_2 + y_1\,(\alpha_1^{\text{new}} - \alpha_1^{\text{old}})\,K(\mathbf{x}_1, \mathbf{x}_2) + y_2\,(\alpha_2^{\text{new,clipped}} - \alpha_2^{\text{old}})\,K(\mathbf{x}_2, \mathbf{x}_2) + b^{\text{old}} \tag{3.52}$$

When both b_1 and b_2 are valid, they are equal. When both new Lagrange multipliers are at bound and if L is not equal to H, then the interval between b_1 and b_2 are all thresholds that are consistent with the Karush-Kuhn-Tucker conditions which are necessary and sufficient conditions for an optimal point of a positive definite QP problem. In this case, SMO chooses the threshold to be halfway between b_1 and b_2 [17].

3.4. Multi-class Classification

The discussion above deals with binary classification where the class labels can take only two values: 1 and −1. In the real world problem, however, we find more than two classes for examples: in fault diagnosis of rotating machineries there are several fault classes such as mechanical unbalance, misalignment and bearing faults. Therefore, in this section the multi-class classification strategy will be discussed.

3.4.1. One-Against-All (OAA)

The earliest used implementation for SVM multi-class classification is one-against-all methods. It constructs k SVM models where k is the number of classes. The ith SVM is trained with all of examples in the ith class with positive labels, and all the other examples with negative labels. Thus given l training data (x_1, y_1), …, (x_l, y_l), where $x_i \in R^n$, i = 1, …, l. and $y_i \in \{1, …, k\}$ is the class of x_i, the ith SVM solve the following problem

$$\text{minimize:} \quad \frac{1}{2}\|\mathbf{w}^i\|^2 + C\sum_{i=1}^{l}\xi_j^i(\mathbf{w}^i)^T \tag{3.53}$$

$$\text{subject to:} \quad (\mathbf{w}^i)^T\phi(\mathbf{x}_j) + b^i \geq 1 - \xi_j^i, \ \text{ if } y = i \tag{3.54}$$

$$(\mathbf{w}^i)^T\phi(\mathbf{x}_j) + b^i \leq -1 + \xi_j^i, \ \text{ if } y \neq i \tag{3.55}$$

$$\xi_J^i \geq 0, \quad j = 1,...,l \tag{3.56}$$

where the training data $\mathbf{x}_i$ is mapped to a higher dimensional space by function ϕ and C is the penalty parameter.

Minimizing Eq. (3.53) means we would like to maximize $2/\|\mathbf{w}_i\|$, the margin between two groups of data. When data is not separable, there is a penalty term $C\sum_{i=1}^{l}\xi_{i,i}$ which can reduce the number of training errors.

3.4.2. One-Against-One (OAO)

Another major method is called one-against-one method. This method constructs $k(k-1)/2$ classifiers where each one is trained on data from two classes. For training data from the ith and the jth classes, we solve the following binary classification problem.

$$\text{minimize:} \quad \frac{1}{2}\|\mathbf{w}^{ij}\|^2 + C\sum_{t}\xi_t^{ij}(\mathbf{w}^{ij})^T \tag{3.57}$$

$$\text{subject to:} \quad (\mathbf{w}^{ij})^T\phi(\mathbf{x}_t) + b^{ij} \geq 1 - \xi_t^{ij}, \ \text{ if } y_t = i \tag{3.58}$$

$$(\mathbf{w}^{ij})^T \phi(\mathbf{x}_t) + b^{ij} \leq -1 + \xi_t^{ij}, \;\; \text{if } y_t = j \tag{3.59}$$

$$\xi_t^{ij} \geq 0, \quad j = 1, \ldots, l \tag{3.60}$$

There are different methods for doing the future testing after all $k(k-1)/2$ classifiers are constructed. After some tests, the decision is made using the following strategy: if sign $((\mathbf{w}^{ij})^T\phi(\mathbf{x})+b^{ij})$ says $\mathbf{x}$ is in the ith class, then the vote for the ith class is added by one. Otherwise, the jth is increased by one. Then $\mathbf{x}$ is predicted in the class using the largest vote. The voting approach described above is also called as Max Win strategy.

3.4.3. Direct Acyclic Graph (DAG)

In this method, the training process is similar to OAO strategy by solving $k(k-1)/2$ binary SVM. However, in the testing process, it uses a rooted binary directed acyclic graph which has $k(k-1)/2$ internal nodes and k leaves. Each node is binary SVM of ith and jth classes. Given a test samples $\mathbf{x}$, starting at the root node, the binary decision function is evaluated. Then it moves to either left or right depending on the output value. The detail explanation of this method is suggested to see reference [18].

4. Wavelet Support Vector Machine (W-SVM)

The idea of wavelet analysis is to approach a function or signal using a family of functions which are produced by translation and dilatation of the mother wavelet function $\psi_{a,b}(x)$

$$\psi_{a,b}(x) = |a|^{-1/2}\,\psi\left(\frac{x-b}{a}\right) \tag{3.61}$$

where x, a, $b \in R$, a is the dilatation factor and b is the translation factor. The wavelet transform of any function $f(x)$ can be expressed as

$$W_{a,b}(f) = \langle f(x), \psi_{a,b}(x) \rangle,\ \ f(x) \in L_2(R) \tag{3.62}$$

where the notation $\langle , \rangle$ refers to inner product in $L_2(R)$.

Eq. (3.62) means that any function $f(x)$ can be decomposed on wavelet basis $\psi_{a,b}(x)$ if it satisfies the condition [19,20]

$$C_\psi = \int_0^\infty \frac{|H(\omega)|^2}{|\omega|} d\omega < \infty \tag{3.63}$$

where $H(\omega)$ is Fourier transform of $\psi_{a,b}(x)$.

Following [19], the function $f(x)$ can be reconstructed as follows

$$f(x) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_0^\infty W_{a,b}(f)\ \psi_{a,b}(x)\ \frac{da}{a^2} db \tag{3.64}$$

To approximate Eq. (3.64), then the finite can be written as

$$\hat{f}(x) = \sum_{i=1}^{l} W_i \psi_{a_i,b_i}(x) \tag{3.65}$$

Using Eq. (3.65), $f(x)$ can eventually be approximated by $\hat{f}(x)$.

For a common multidimensional wavelet function, the mother wavelet can be given as the product of one-dimensional (1-D) wavelet function [20]

$$\psi(x) = \prod_{i=1}^{N} \psi(x_i) \tag{3.66}$$

where $\mathbf{x} = (x_1, ..., x_N) \in R^N$. So, every 1-D wavelet mother $\psi(x)$ must satisfy Eq. (3.63).

Recalling the decision function for SVM in Eq. (3.43), the dot product can be replaced using kernel function as it was done by [11], so that $K(\mathbf{x}, \mathbf{x}') = K(\langle \mathbf{x} \cdot \mathbf{x}' \rangle)$. In SVM theory, any function which satisfies the Mercer's condition can serve as kernel function [11,14].

Suppose K is a continuous symmetric function on R^N, such that integral operator T_K: $L_2(R^N) \rightarrow L_2(R^N)$,

$$(T_K) f(\cdot) = \int_{R^d} K(\cdot,\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \tag{3.67}$$

is positive. Let $\phi_i \in L_2(R^N)$ be the eigenfunction of T_k associated with the eigenvalue $\lambda_i \geq 0$ and be normalized in such a way that $\| \phi_i \|_{L_2} = 1$, then the kernel function $K(\mathbf{x},\mathbf{x}')$ can be expanded as

$$K(\mathbf{x},\mathbf{x}') = \sum_{i=1}^{\infty} \lambda_i \, \phi_i(\mathbf{x}) \, \phi_i(\mathbf{x}') \tag{3.68}$$

and must satisfy the positivity condition of the following integral [14]

$$\int \int_{L_2 \otimes L_2} K(\mathbf{x},\mathbf{x}') f(\mathbf{x}) f(\mathbf{x}') \, d\mathbf{x} \, d\mathbf{x}' \geq 0 \ , \forall \ f \in L_2(R^N) \tag{3.69}$$

For building a new kernel using wavelet, it may be helpful to refer to the frame theory, introduced by Duffin and Schaeffer [21], which is an extension of the normalized orthogonal basis. In the frame theory, one can reconstruct perfectly a function *f* in a Hilbert space *H* from its inner product $\langle , \rangle$ with family vectors $\{\psi_k\}$ if they satisfy

$$A \| f \|^2 \leq \sum_k |\langle f, \overline{\psi}_k \rangle|^2 \leq B \| f \|^2 \tag{3.70}$$

where the constants A and B satisfy the condition $0 < A \leq B < \infty$.

Any function in Hibert space can be decomposed as follows

$$f = \sum_k \langle f, \overline{\psi}_k \rangle \psi_k = \sum_k \langle f, \psi_k \rangle \overline{\psi}_k \tag{3.71}$$

where $\overline{\psi}_k = (T^* T)^{-1} \psi_k$ is the dual frame of ψ_k and *T* is the frame operator [12, 31].

In $L_2(R^N)$, if $f = \{\psi_i\}$ is a frame and $\{\lambda_i\}$ is a positive increasing sequence, a function $K(\mathbf{x},\mathbf{x}')$ can be given by

$$K(\mathbf{x},\mathbf{x}') = \sum_{i=1}^{\infty} \lambda_i \, \psi_i(\mathbf{x}) \, \psi_i(\mathbf{x}') \tag{3.72}$$

Eq. (3.72) is similar to Eq. (3.68) since both of them satisfy the condition for kernel function. Moreover, a mother wavelet $\psi_{a,b}(x)$ is called a frame wavelet if $\psi \in L_2(R^N)$, $a > 1$, $b > 0$ and the family function

$$\{\psi_{mn}\} = \{D_{am} T_{nb} \psi\} \tag{3.73}$$

where D and T are unitary dilatation operator and unitary translation operator, respectively, while a is scale parameter and b is translation parameter.

A wavelet kernel function can be constructed by any mother wavelet which can generate frame wavelet while satisfying the Mercer's condition in Eq. (3.69). In addition to the inner product, there exists a kernel called translation–invariant kernel [22, 23] such that

$$K(\mathbf{x},\mathbf{x}') = K(\langle \mathbf{x} - \mathbf{x}' \rangle) \tag{3.74}$$

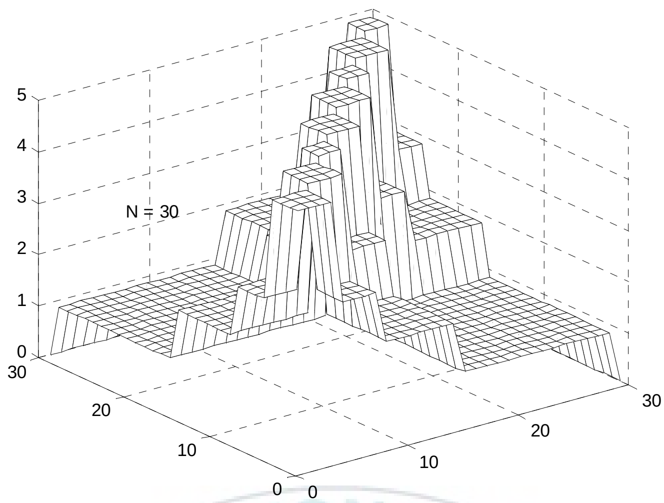
If the translation–invariant kernel is admissible in SVM kernel function, then the necessary and sufficient condition of Mercer's theorem must be satisfied. The other theorem stated that a translation-invariant kernel is an admissible support vector (SV) kernel if only if the following Fourier transforms [22]

$$F[K](\omega) = (2\pi)^{-N/2} \int_{R^N} \exp(-j(\omega \cdot \mathbf{x}))\, K(\mathbf{x})\, d\mathbf{x} \tag{3.75}$$

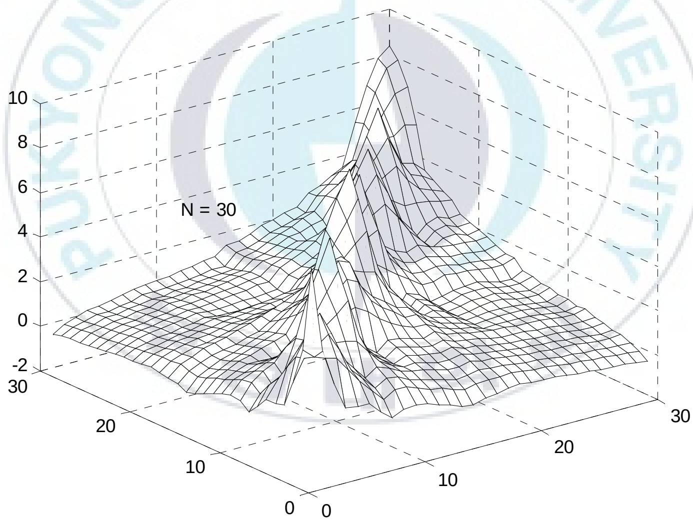
is non-negative. Based on the mother wavelet, the wavelet kernel which satisfies the translation– invariant theorem can be given as

$$K(\mathbf{x},\mathbf{x}') = K(\mathbf{x} - \mathbf{x}') = \prod_{i=1}^{N} \psi\left(\frac{x_i - x_i'}{a_i}\right) \tag{3.76}$$

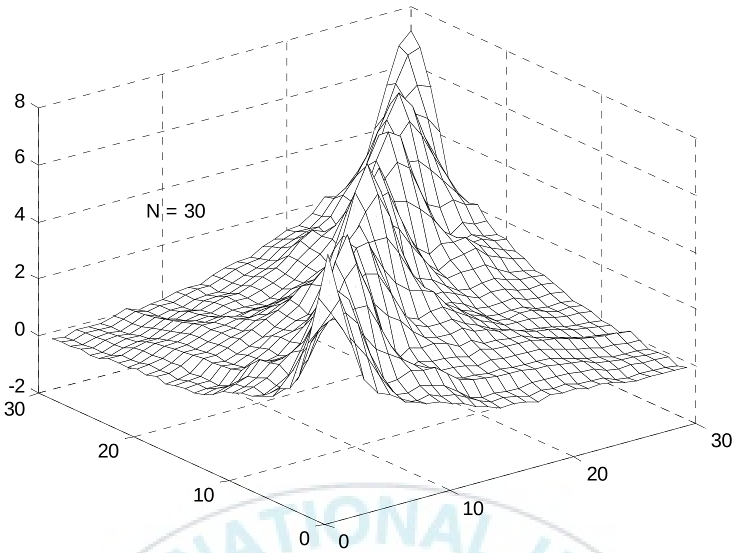
The construction of wavelet kernel function using Haar, Daubechies, and Symmlet can be shown in Fig. 3.2.



(a) Haar kernel



(b) Daubechies kernel



(c) Symlet kernel

Fig. 3.2. Wavelet kernel function.

References

[1] Turk M, Pentland A, Face recognition using eigenfaces [C], Proceeding IEEE Conference on Computer Vision and Pattern Recognition, 1991, pp. 586-591.

[2] D. Vijayaseenan, S. Madhvanath, Ramakrishnan A.G., Principal component analysis for online handwritten character recognition, 17 International Conference paatern Recognition, Cambridge, United Kingdom, 2004, pp. 327-330.

[3] Pinkowski, B., J. Finnegan-Green, Principal component analysis of speech spectrogram images, Pattern Recognition 30 (1996) 777-787.

[4] I.J. Jolliffe. Principal Component Analysis. New York: Springer; 1986.

[5] J.F. Cardoso, Blind signal separation: statistical principles, Proceeding of the IEEE 86(10), 1998, pp. 2009-2020.

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

[7] A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis, IEEE Trans. Neural Networks 10 (1999) 626-634.

[8] B. Scholköpf, A. Smola, K.R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Computation 10(5) (1998) 1299-1319.

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

[10] A. Cichocki, S. Amari, K. Siwek, T. Tanaka, ICALAB; Toolbox for ICA, BSS and BSE, http://www.bsp.brain.riken.jp/ICALAB/ICALABSignalProc/.

[11] V.N. Vapnik, The Nature of Statistical Learning Theory, New York: Springer; 1995.

[12] B. Boser, I. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers, Proceedings of Fifth Annual Workshop on Computational Learning Theory, New York, 1992.

[13] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20(3) (1995) 273-297.

[14] N. Cristianini N.J. Shawe-Taylor, An Introduction to Support Vector Machines, Cambridge: Cambridge University Press, 2000.

[15] V.N. Vapnik. Estimation Dependences Based on Empirical Data, Berlin: Springer Verlag, 1982.

[16] E. Osuna, R.R. Freund & F.F. Girosi, Improved training algorithm for support vector machines, Proceeding of IEEE Neural Networks for Signal Processing, 1997, pp. 276-285.

[17] J. Platt, Fast training of support vector machines using sequential minimal optimization, in: B. Scholkopf, et al. (Eds.), Advances in Kernel Methods–

Support Vector Learning. Cambridge: MIT Press, 1999, pp. 185-208.

[18] C.W. Hsu, C.J. Lin, A comparison of methods for multi-class support vector machines, IEEE Trans. Neural Networks 13 (2002) 415-425.

[19] I. Daubechies, The wavelet transform, time-frequency localization and signal analysis, IEEE Transaction on Information Theory 36 (1990) 961-1005.

[20] Q.H. Zhang, A. Benveniste, Wavelet networks. IEEE Transaction on Neural Networks 3 (1992) 889-898.

[21] R.J. Duffin, A. Schaeffer, A class of nonharmonic Fourier series. Transaction of American Mathematics Society 72 (1952) 341-366.

[22] A. Smola, B. Scholköpf, K.R. Müller, The connection between regularization operators and support vector kernels, Neural Network 11 (1998) 637-649.

[23] L. Zhang, W. Zhou, L. Jiao, Wavelet support vector machine, IEEE Transaction on System, Man, and Cybernetics-Part B: Cybernetics 34 (1) (2004) 34-39.

IV. SVM Based Fault Diagnosis for Induction Motors

1. Introduction

Three phase induction motors are the motors most frequently used in industry. They are simple, rugged, relatively low-price, and easy to maintain. In this chapter, the basic principle of three phase induction motors is reviewed including its general structure and construction. Moreover, fault in induction motors that are frequently occurred and measurement for fault diagnosis will be reviewed.

Fault diagnosis of induction motors is also presented which is the main part of this chapter. First, the existed method for fault diagnosis of induction motor is reviewed and then followed by introducing the proposed method. Finally, case study of fault diagnosis of induction motor is presented based on vibration and current signals.

2. Structure and Operation

A three-phase induction motor, presented in Fig. 4.1 has two main parts: a stationary stator and a revolving rotor. The rotor is separated from the stator by a small air gap that ranges from 0.4 mm to 4 mm, depending on the power of motor.

The stator consists of a steel frame that supports a hollow cylindrical core made up of stacked laminations. A number of evenly spaced slots, punched out of the internal circumference of the laminations, provide the space for the stator winding.

The rotor is also composed of punched laminations. These are carefully stacked to create a series of rotor slots to provide space for the rotor winding. There are two types of rotor windings: conventional 3-phase windings made of insulated wire and squirrel-cage windings. The type of winding give rise two main classes of motors: squirrel-cage induction motors and wound-rotor induction motors.

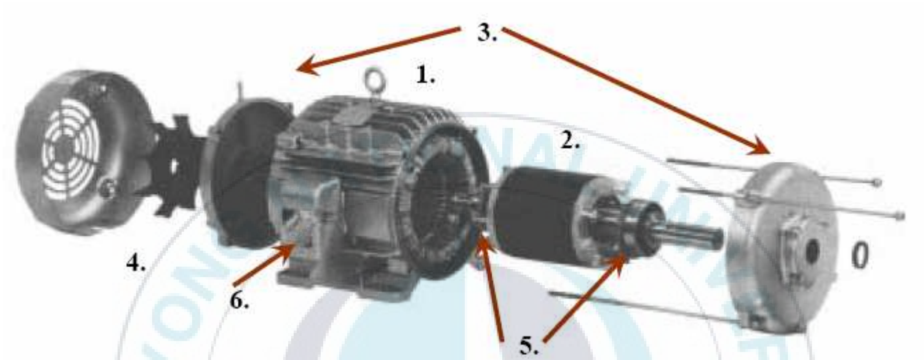


Fig. 4.1 Exploded view of cage motor: Stator (1), Rotor (2) End-caps (3), Cooling fan (4), Ball bearings (5), Terminal box (6) [1].

A squirrel-cage rotor is composed of bare per bars, slightly longer than rotor, which are pushed into the slots. The opposite ends are welded to two copper end-rings, so that all the bars short-circuited together. The entire construction resembles a squirrel-cage, from which the name is derived. In small and medium size of motors, the bars and end-rings are made of die-cast aluminum, molded to form an integral block

Another type is a wound-rotor has a 3-phase winding, similar to the one of the stator. The winding is uniformly distributed in the slot and is usually connected in 3-wire *wye*. This motor is, however, less efficient than the squirrel-cage induction motor, and it is used only when a squirrel-cage induction motor cannot deliver the high enough starting torque.

When the stator winding of a three-phase induction motor is connected to a three-phase power source, it produces a magnetic field that is a constant in magnitude and revolves around the rotor at the synchronous speed. If f is the frequency of the current in the stator winding and P is the number of poles, the synchronous speed of the revolving field is

$$n_s = \frac{120f}{P} \tag{4.1}$$

where n_s is synchronous speed (r/min), f is frequency of the source (Hz) and P is number of poles. This equation shows that the synchronous speed increases with frequency and decreases with number of poles.

The revolving field induces electromotive force (EMF) in the rotor winding. Since the rotor winding forms a closed loop, the induced EMF in each coil gives rise to an induced current in that coil. When a current-carrying coil is in a magnetic field, it experiences a force that tends to rotate it. The rotor receives its power by induction only when there is a relative motion between the rotor speed and the revolving field. Since the rotor rotates at a speed lower than the synchronous speed of the revolving field, an induction motor is also called an asynchronous motor.

The slip of induction motor s, is defined as the difference between the synchronous speed and the rotor speed, expressed as a percent (or per unit) of synchronous speed. The per unit slip is given by equation

$$s = \frac{n_s - n}{n_s} \tag{4.2}$$

where n is rotor speed (r/min), The slip s is practically zero at no-load and is equal to 1 (or 100%) when rotor is locked.

3. Fault Occurrence and Measurement for Diagnosis

The faults frequently occurred in induction motors components are rotor, stator and bearing defects. Based on EPRI which has conducted large survey on motors fault of 5000 sample motors, 97% among them are three-phase squirrel-cage induction motors. The fault occurrence based on the survey is presented in Fig. 4.2. Most common fault is worn bearing that generate excessive vibration, noise and possible misalignment of the rotor shaft. Most of the stator related faults are due to degraded insulation in stator windings causing an inter-turn, phase-to-phase or phase-to-ground short circuits. Other case is rotor fault which can be divided into faults related to motor eccentricity and physical damage of the rotor and they are usually slowly although in the end the broken bars may damage the stator windings.

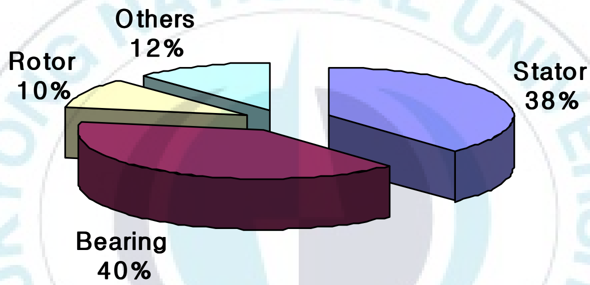


Fig. 4.2 EPRI survey on occurrence of motor faults [2].

It is found out that a variety of measurements can be applied to collect information that is useful in the detection of induction motor faults. In this dissertation, two of them are elaborated using stator current of the motor and vibrations of the motor. Vibration analysis has been used in motor fault detection for decades. Each fault in a rotating machine produces vibrations with distinctive characteristics that can be measured and compared with reference ones in order to perform the fault detection and diagnosis. Motor current monitoring is also called motor current signature analysis (MCSA) and it is widely studied, because no extra instrumentation is needed, if the faults can be detected based on the current. It is also claimed that MCSA give the same information on motor condition as vibration measurements [3].

In this dissertation, the faults frequently occurred in induction motors are reviewed as follows:

3.1. Bearing Fault

A bearing consists of two rings inner and outer, between which a set of balls or rollers rotate in raceways. Fig. 4.3 shows the part of a deep groove ball bearing. Under normal operating conditions of balanced load and good alignment, fatigue failure begins with a small fissure, located between the surface of the raceway and rolling elements, which gradually propagate to the surface generating detectable vibrations and increasing noise levels [4]. Continued stressing causes fragments of the material to break loose producing a localized fatigue phenomenon known as flaking or spalling [5]. Once started, the affected area expands rapidly contaminating the lubrication and causing localized overloading over the entire circumference of the raceway.

Eventually, the failure results in rough running of the bearing. While this is the normal mode of failure in rolling element bearings, there are many other conditions which reduce time of bearing failure. These external sources include contamination, corrosion, improper lubrication, improper installation or brinelling.

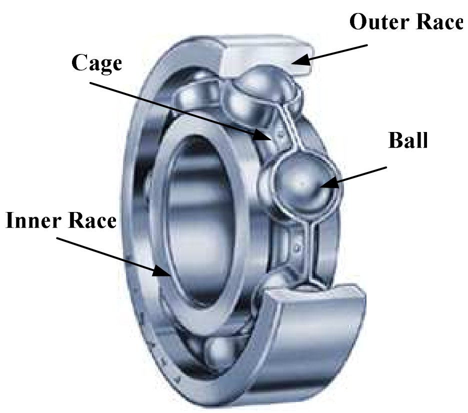


Fig. 4.3 The structure of a deep groove ball bearing.

Contamination and corrosion frequently accelerate bearing failure because of the harsh environments present in most industrial settings. Dirt and other foreign matter that is commonly present often contaminate the bearing lubrication. The abrasive nature of these minute particles, whose hardness can vary from relatively soft the diamond like, causes pitting and sanding actions that give way to measurable wear of the balls and raceways [5]. Bearing corrosion is produced by the presence of water, acids, deteriorated lubrication and even perspiration from careless handling during installations [4,5]. Once the chemical reaction has advanced sufficiently, particles are worn off resulting in the same abrasive action produced by bearing contamination. Improper lubrication includes both under and over lubrication. In either case, the rolling elements are not allowed to rotate on the designed oil film causing increased levels of heating. The excessive heating causes the grease to break down which reduces its ability to lubricate the bearing elements and accelerates the failure process. When the lubrication conditions become inadequate, the increased friction results in metal – metal contact.

Installation problems are often caused by improperly forcing the bearing onto the shaft or in the housing. This produces physical damage in the form of brinelling or false brinelling of the raceways which leads to premature failure. Misalignment of the bearing, which occurs in the four ways depicted in Fig. 4.4, is also a common result of defective bearing installation. The most common of these is caused by tilted races [5].

Brinelling is the formation of indentations in the raceways as a result of deformation caused by static overloading. While this form of damage is rare, a form of “false brinelling” occurs more often. In this case, the bearing is exposed to vibrations while even though lightly loaded bearings are less susceptible, false brinelling still happens and has even occurred during the transportation of uninstalled bearings [4].

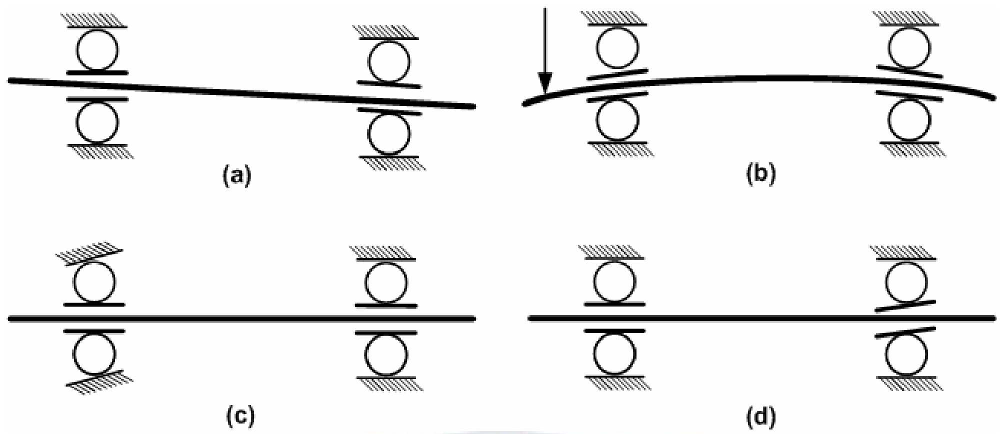


Fig. 4.4 (a) Misalignment (Out-of-Line), (b) Shaft deflection, (c) Crooked or tilted outer race (d) crooked or tilted inner race

Regardless of the failure mechanism, defective rolling element bearings generate mechanical vibrations at the rotational speeds of each component. These characteristic frequencies, which are related to the raceways and the balls or rollers, can be calculated from the bearing dimensions and the rotational speed of the machine. Mechanical vibration analysis techniques are commonly used to monitor these frequencies in order to determine the condition of the bearing.

The characteristic frequencies of bearing are as follow

$$BPFO = (N/2)\, f_r \{1 - (B/P) \cos\phi\} \tag{4.3}$$

$$BPFI = (N/2)\, f_r \{1 + (B/P) \cos\phi\} \tag{4.4}$$

$$BSF = (P/2B)\, f_r \{1 - (B/P)^2 \cos^2\phi\} \tag{4.5}$$

$$FTF = (f_r /2) \{1 - (B/P) \cos\phi\} \tag{4.6}$$

BPFO is ball pass frequency of the outer race; generated by rollers passing over defective outer race. *BPFI* is ball pass frequency of the inner race; generated by rollers passing over defective inner race. *BSF* is ball spin frequency; generated by ball defects. *FTF* is fundamental train frequency; generated by cage defects or improper movements. Then, N is number of rolling elements, P is pitch diameter (mm), B is ball or roller diameter (mm) and f_r is rotating speed in revolution per second.

The frequencies in Eqs. (4.3)-(4.6) are valid for ideal bearings; in practice, the rolling element slides in addition to its rotation. Using a sliding factor that ranges from 0.8-1.0, this phenomena can be taken account. In both literature and practice the equations are often replaced by approximate equation [6] which can be used when the exact bearing geometry is not known. A characteristic frequency using approximate formula for outer race and inner race defects are

$$f_o = 0.4Nf_r \tag{4.7}$$

$$f_i = 0.6Nf_r \tag{4.8}$$

Schoen [7] implemented motor current in technique to detect rolling-element bearing fault in induction motors. Line current spectral components are predicted at frequencies of

$$f_{bng} = | f_e \pm mf_v | \tag{4.9}$$

where f_v is one of the characteristic vibration frequencies, f_e is the supply frequency, and m = 1, 2, 3, … . Although the magnitudes of this harmonic component are small compared to other spectral constituents, they fall at different location from those of the supply and machine inherent slot harmonics. This phenomenon makes it feasible to distinguish between healthy and faulty operations.

3.2. Stator Fault

Stator winding faults constitute almost 30-40% on induction motor faults according the survey. These faults are usually short circuit between a phase winding and the ground or between two phases. It is strongly believed that such fault initiate as undetected turn-to-turn faults that develop to a major short circuit.

Stator winding fault might have a destructive effect of the stator coils.

Armature of stator insulation can fail due to several reasons as follows:

1. Short circuit or starting stress.
2. Stack core lamination, slot wedges and joints.
3. Electrical discharge.
4. High stator core or winding temperature.
5. Loose bracing for end winding.
6. Contamination due to oil, moisture and dirt.
7. Leakage in cooling system.

There are several methods proposed to detect the mentioned faults. Cash [8] summed up the machine line-to-neutral voltages instantaneously and filtered out the undesired saturation, slots and other sound operation harmonic. The RMS value of the remaining voltage component was utilized to detect the existence and severity of stator inter-turn faults, the standard deviation of the RMS line current of an induction motor was used to detect stator inter-turns [9].

Penman [10] monitored the axial leakage flux resulting from the stator winding to detect and locate stator inter-turns. The voltage induced in a search coil wound concentrically around the machine shaft was proportional to this flux component. Some spectral constituents of this voltage were observed to detect a turn-to-turn fault. These frequencies are given by

$$f_t = \left[k \pm \left(\frac{n}{p} \right)(1-s) \right] f_e \tag{4.10}$$

where k = 1, 3 and n = 1, 2, 3, …, (2p-1), p is the number of pole pairs, s is the slip and f_e is the supply frequency. The location of the inter-turn fault could be specified using four auxiliary winding mounted symmetrically in the four quadrants of the motor near the end winding. The flux RMS magnitudes at the various locations were measured. The change in readings from the four coils could be used to triangulate the area of the unbalanced flux, and hence, locate the shorted turn.

According to the modes of stator winding failure, there are five types of modes, which are illustrated in Fig. 4.5.

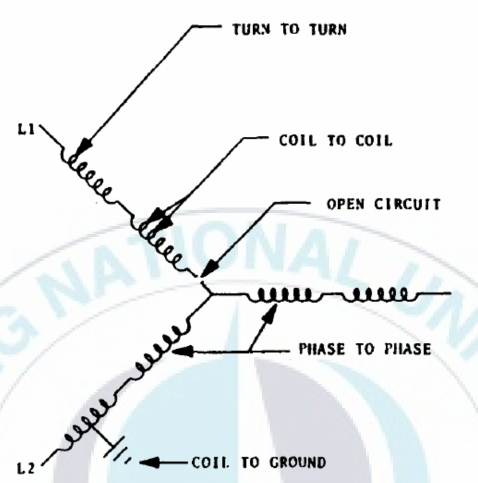


Fig. 4.5 A possible failure modes in *wye*-connected stator winding [11].

Bonnet [11] reported in detail cause and analysis of stator faults those are influenced by various of stresses such as the following:

3.2.1. Thermal Stress

The stress in induction motor that caused by temperature effects such as thermal aging and thermal overloading. The AIEE 510 and IEEE 275 test procedures can be used to determine the effect of temperature on the winding insulation system. Thermal overloading is influenced by various factors i.e. voltage variations, unbalanced phase voltage, cycling, overloading, obstructed ventilation and ambient temperature. The relationship between the various classes of insulation and operating temperature is presented in Fig. 4.6.

Unless the operating temperature is extremely high, the normal effect of the thermal aging is to render the insulation system vulnerable to other influencing factor or stresses that actually produce the failure. The detail of this information is reported in aforementioned reference.

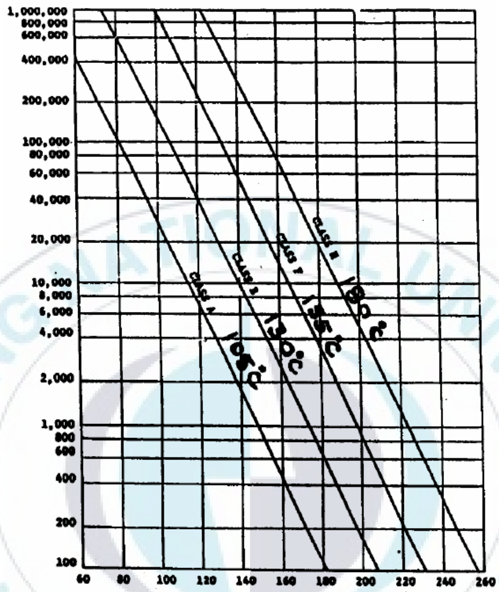


Fig. 4.6 Total winding temperature (°C) versus life [11].

3.2.2. Electrical Stress

Electrical stress are generally discussed as failures in the windings such as phase-to-phase, turn-to-turn, or phase-to-ground shorts. Testing to determine the integrity of the insulation is paramount to long motor life. Checking the integrity of the insulation can be accomplished by the MCE standard test. Insulation can also have tracking occur in which the insulation develops a small hole which leakage to ground. If the motor is contaminated with conductive foreign materials, this will create a path to ground causing the insulation to burn, which causes further deterioration of the insulation. Keeping the insulation dry and contaminant free will help to minimize or prevent tracking from occurring. Another method employed to prevent tracking is to use insulation capable of being completely immersed in accordance with NEMA MG 1-20.49 and IEEE 429.

3.2.3. Mechanical Stress

The stator coils can and do move during operation of the motor, especially when the motor is started. When the motor is started, the current in the coils is at highest which result in a high magnetic force that causes the coils to vibrate at two times line frequency. This vibration causes the coils to move, which can result in damage to stator, rotor and other motor components. Bearing failures and misalignment can cause the rotor to strike the stator, which can result in grounded coils, excessive heat generation, or severe damage to both the rotor and stator.

Some of the common causes of the winding failures, which can fit into the miscellaneous mechanical type of failure, are as follows:

1. Rotor balancing weights coming loose and striking the stator.
2. Rotor fan blades coming loose and striking the stator.
3. Loose nuts and bolts striking the stator.
4. Foreign particles entering the motor through the ventilation system and striking the stator.
5. A defective rotor (usually open rotor bars), causing the stator to overheat and fails.
6. Poor lead lugging of connections from the motor leads to the incoming line leads, causing overheating and failures.
7. Broken lamination teeth striking the stator due to fatigue.

3.2.4. Environmental Stress

The quickest way to discuss environmental stress is to call it what it is: contamination. Contamination is anything in the motor that is not supposed to be there. Any foreign material that enters the motor can cause environmental stress. Some examples are moisture, oil, dirt, coal, dust, etc. All of these contaminants can have the following effects on the motor:

1. Reduction in heat dissipation, which will increase operating temperature, thereby reducing insulation life.
2. Premature bearing failure due to high localized stresses.
3. Breakdown of the insulation system, causing shorts and grounds.

3.3. Rotor Fault

The reasons for rotor bars and end-ring breakage are several. They can be caused by

1. Thermal stress due to overload and hotspot or excessive looses and sparking.
2. Magnetic stresses caused by electromagnetic forces, unbalanced magnetic pull, electromagnetic noise and vibration.
3. Residual stresses due to manufacturing imperfections.
4. Dynamic stresses arising from shaft torque, centrifugal forces and cyclic stresses.
5. Environmental stresses caused by for example contamination and abrasion of rotor material due to chemical or moisture.
6. Mechanical stresses due to loose laminations, fatigued parts, bearing failures, etc.

Motor current signature analysis was extensively used to detect broken rotor bar and end ring faults in induction motors [12,13]. The sideband components used to detect broken rotor bars is given by

$$f_b = (1 \pm 2s) f_e \tag{4.11}$$

while the lower sideband was fault related and the upper sideband was due to consequent speed oscillation. Bellini [14] stated the summation of magnitudes of these two sideband components was a good diagnostics index. It was concluded that MCSA was superior to signature analysis of current space vector modulus and instantaneous power and torque. The actual sequence of sidebands was given by [15]

$$f_b = (1 \pm 2ks) f_e \tag{4.12}$$

where k = 1, 2, 3, … .

Considering the speed ripple effects, it was reported that other frequency components, which could be observed in the stator current spectrum, are given by

$$f_b = \left[\left(\frac{k}{p}\right)(1-s) \pm s\right] f_e \tag{4.13}$$

where p is the number of pole pairs, and k = 1, 2, 3, … .

The other method for rotor fault detection is reported using current Park's vector approach to diagnose rotor cage faults of three-phase induction motors [16]. This technique can be used to distinguish between the effect of this fault and that associated with driving time-varying loads. Rotor cage faults can be detected by the identification of an elliptic figure in Park's vector representation. When the load has low-frequency oscillating component, the current Park's vector pattern is an ellipse oriented along the first quadrant of the coordinate axes. In the presence of rotor cage fault, the pattern of ellipse becomes oriented in the second quadrant of the coordinate axes.

3.4. Eccentricity

Rotor eccentricity, which results in uniform airgap, is divided into two catagpries, static and dynamic. In static eccentricity case, the airgap has a fixed minimal position, whereas this position rotates with the rotor in case of dynamic eccentricity. In practice, both of types occur simultaneously. Due to some designs and manufacturing imperfections, up to 10% eccentricity is allowed. Higher order of eccentricity can cause rotor-to-stator rub, resulting in damage of rotor and/or stator winding or core.

Eccentricity faults could be diagnosed by monitoring the airgap flux in induction motors. Internal and external search coils were placed in the stator and the spectral constituents of their induced voltage were observed for diagnosing component at

$$f_{ec} = f_e \pm f_r \tag{4.14}$$

where f_e is supply frequency and f_r is the rotational frequency.

Dorrel [17] monitored casing vibration components at a frequency $2f_e \pm f_r$ to diagnose eccentricity faults in induction motors. Motor current signature analysis (MCSA) was used extensively to diagnose eccentricity faults in three-phase induction motors. Specific frequencies related to fault are given by

$$f_{ec} = \left[(kR \pm n_d)\frac{(1-s)}{p} \pm v \right] f_e \tag{4.15}$$

Where k is any positive integer, R is the number of rotor bars, p is the number of pole pairs, n_d is the eccentricity order (n_d = 0 for static eccentricity, n_d = 1 for dynamic eccentricity), s is the motor slip, v is the order of some harmonics present in the power supply driving the motor (v = 1, 3, 5, …).

In the case of static eccentricity, principal slot harmonic and supply time harmonics contribute to these components. If the order of one of this harmonics is a multiple of three, it may not theoretically appear in the spectrum of a balanced machine. However, it was shown that for a specific combination of the number of fundamental pole pairs and number of rotor slots, the machine would give rise to only static or only dynamic eccentricity related components [18].

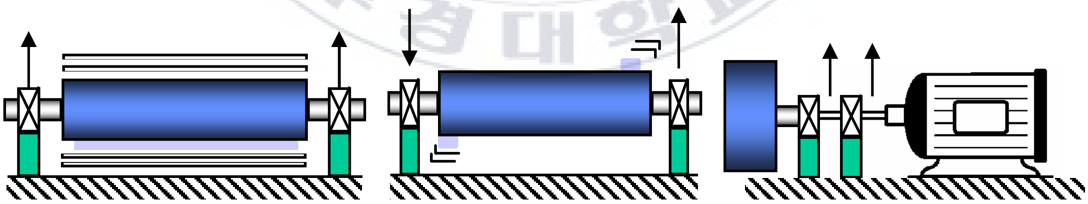
Obaid [19] used MCSA to diagnose eccentricity faults in three-phase induction motors by observing the components

$$f_{ec} = \left[\left(\frac{m}{p} \right) (1 - s) \pm 1 \right] f_e \tag{4.16}$$

where m is a positive integer. The RMS value of each component was calculated after filtering out the fundamental. The RMS values were compared to a preset threshold that was determined the observation of sound operation. Under load imbalance, and horizontal and vertical misalignment conditions, the machine gave rise to such harmonic components with magnitudes dependent on the condition.

3.5. Unbalance Mass

Mass unbalance is the most common fault associated with rotating shaft. It occurs when the geometric center (shaft centerline) and the mass center of a rotor do not coincide. There are three types of unbalance (Fig. 4.7): Static unbalance coupled unbalance and overhung rotor unbalance. Static unbalance has equal phase on each bearing, so vibration along with radial direction in phase. While in coupled unbalance, phase changes 180° across bearing, so vibration along with radial direction out phase. Overhung rotor unbalance contains both radial and horizontal vibration, so both static and dynamic unbalance can be seen together.



(a) Static unbalance (b) Coupled unbalance (c) Overhung rotor unbalance

Fig. 4.7 Mass unbalance.

3.6. Bowed Rotor

A bowed rotor or bent shaft usually causes a preload on the bearings. The center of the mass of a bent shaft can be moved far enough away from the geometric center to cause some mass unbalance (Fig. 4.8). A bent shaft is looking like a misalignment in the spectrum. A phase measurement for axial vibration across the shaft will distinguish between misalignment and bent shaft, as the bent shaft will produce a 180 degrees phase shift. Also the vibration style of a bent shaft contains axial and radial direction. Among them, 180° phase shift in axial vibration, while 0° phase shift in radial vibration.

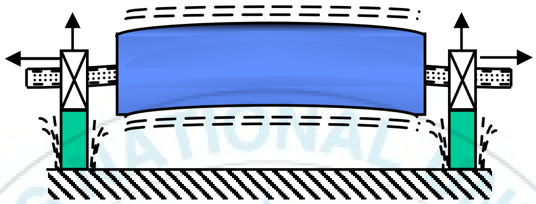


Fig. 4.8 Bowed rotor.

4. Condition Monitoring and Fault Diagnosis of Induction Motors

Induction motor is an essential component in many industrial processes which deals with moving and lifting products. Special attention is urgently required in condition monitoring of induction motors in order to guarantee its stable and high performance. By applying early fault diagnosis of operating induction motors which give incipient fault condition, little effort to overcome such fault can avoid more serious conditions.

Condition monitoring and fault diagnosis methods to identify the faults may involve different types of techniques. These techniques can be described as follows:

Electromagnetic field monitoring, search coils, coils wound around motor shafts (axial flux related detection)

Temperature measurements

Infrared recognition

Radio frequency (RF) emissions monitoring

Noise and vibration monitoring

Chemical analysis

Acoustic noise measurements

Motor current signature analysis

Model, artificial intelligence based techniques

Several methods of condition monitoring and fault diagnosis that related to fault can be detected are presented and compared in Table. 4.1.

Table 4.1 Comparison of detection techniques

Methods	Fault detected				
	Insulating	Stator winding	Air-gap eccentricity	Broken rotor bars	Bearing damage
Vibration	No	No	Yes	Yes	Yes
MCSA	No	Yes	Yes	Yes	Yes
Axial flux	No	Yes	Yes	Yes	No
Lubricating oils debris	No	No	No	No	Yes
Cooling gaps	Yes	Yes	Yes	No	No
Partial discharge	Yes	No	No	No	No

5. The Proposed Fault Diagnosis and Case Studies

In this work, vibration and/or current signature for detection and diagnose of faults in induction motor may be consider as a kind of pattern recognition paradigm. It consists of data acquisition, signal processing, feature extraction and selection-including feature reduction- and faults diagnosis. A novel faults diagnosis method for induction motor is proposed in Fig. 4.9, which is based on feature extraction (linear and nonlinear), the distance evaluation technique and SVM multi-class classification. From Fig. 4.9, the fault diagnosis procedure can be summarized as follows:

Step 1: the data acquisition is carried out and then followed by features calculation using statistical features parameter from time domain and frequency domain.

Step 2: feature extraction is performed by linear and non linear technique via component analysis to reduce the dimensionality. This step is employed to remove the irrelevant features which are redundant and even degrade the performance of the classifier.

Step 3: feature selection is performed using the distance of evaluation technique. This method is chosen due to the simplicity and its reliability.

Step 4: classification process for diagnosing of faults is carried out using SVM based on multi-class classification.

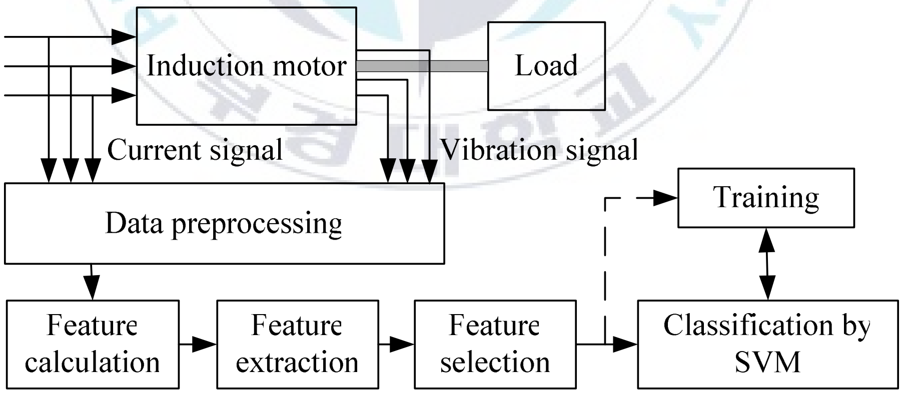


Fig. 4.9 The proposed method for fault diagnosis of induction motor.

In this part, several case studies based on method of feature extraction, signal source, and classification using SVM are presented as follows

5.1. Case Study 1: Using Linear Feature Extraction

5.1.1. Experiment and 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. 4.10. 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. 4.11 and Table 4.2.

Fig. 4.10 Test rig for 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 16,384.

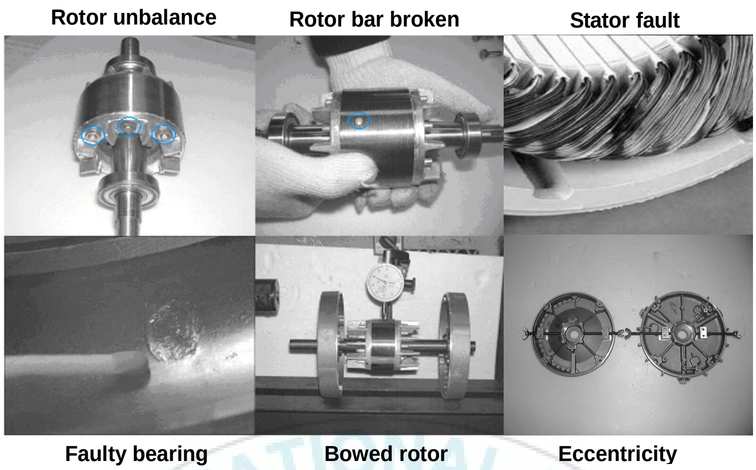


Fig. 4.11 The condition of faulty motor.

Table 4.2 Description of faulty motors

Fault condition	Fault description	Others
Broken rotor bar	No. of broken bar: 12 ea	Total number of 34 bars
Bowed rotor	Maximum bowed shaft deflection: 0.075 mm	Air-gap: 0.25 mm
Faulty bearing	A spalling on outer raceway	#6203
Rotor unbalance	Unbalance mass (8.4 g)on the rotor	
Eccentricity	Parallel and angular misalignments	Adjusting the bearing pedestal
Phase unbalance	Add resistance on one phase	8.4%

5.1.2. 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 3 parameters from frequency domain (rms frequency,

frequency center and root variance frequency) using vibration acceleration signal at the three directions and three–phase current signals. The total of feature parameters can be shown in Table 4.3.

Table 4.3 Feature parameters

Signals	Position	Feature parameters: Time domain	Feature parameters: 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		

5.1.3. Feature Extraction

Basically feature extraction is mapping process of data from higher dimension into low dimension space. This step is intended to avoid the curse of dimensionality phenomenon. ICA and PCA were used to reduce the feature dimensionality that contains 95 % variation of eigenvalue. In this work, feature extraction produced 24 independents components (ICs) and principal component (PCs) based on the eigenvalue. Also, from feature extraction using ICA and PCA, we can understand that there is a change from data features becomes components which are independent and uncorrelated, respectively. The first three independent and principal components are plotted in Figs. 4.12 and 4.13. It can be observed that the clusters for eight conditions are well separated. Nevertheless, the performance of ICA is better than PCA does in clustering of each condition. It can be seen that feature extraction using ICA can separate well almost all of conditions without overlapping except normal and phase unbalance, while PCA produced overlapping in phase unbalance, rotor unbalance and rotor broken bar, also angular misalignment and parallel alignment.

5.1.4. Feature Selection

To select the optimal feature of ICs and PCs that can represent well the condition of induction motors, a feature selection method based on the distance evaluation technique is presented [20,21]. Let that joint feature set of C condition-patterns α_1, α_2, …, α_c are

$$\{q^{(i,k)}, \quad i=1,...,C; \quad k=1,...,N_i\} \tag{4.17}$$

where $q^{(i,k)}$ is the kth feature of α_i, and N_i is the number of feature in α_i.

The average distance of all features in α_i. can be determined as follows

$$D_i = \frac{1}{2}\frac{1}{N_i}\sum_{j=1}^{N_i}\frac{1}{N_i-1}\sum_{k=1}^{N_i}|q^{(i,j)} - q^{(i,k)}| \tag{4.18}$$

The average distance of D_i, i =1, 2, …, C is

$$D_a = \frac{1}{C}\sum_{i=1}^{C} D_i \tag{4.19}$$

Introducing Eq. (4.18) into Eq. (4.19) yields

$$D_a = \frac{1}{C}\sum_{1}^{C}\frac{1}{N_i-1}\sum_{k=1}^{N_i}|q^{(i,k)} - \rho^{(i)}| \tag{4.20}$$

where $\rho^{(i)} = \frac{1}{N_i}\sum_{k=1}^{N_i} q^{(i,k)}$ is the mean of all features in α_i.

The average distance of C different condition-patterns α_1, α_2, …, α_c is

$$D_b = \frac{1}{C}\sum_{i=1}^{C}|\rho^{(i)} - \rho| \tag{4.21}$$

where $\rho = \frac{1}{C}\sum_{i=1}^{C}\frac{1}{N_i}\sum_{k=1}^{N_i} q^{(i,k)}$.

When the average distance D_a inside certain condition-pattern is smaller and the average distance D_b between different condition patterns is bigger, the average represents the optimal features well. The evaluation criteria for optimal features is defined as

$$\delta_A = \frac{D_a}{D_b} \tag{4.22}$$

So, according to the bigger distance evaluation criteria of δ_A, the optimal features can be selected from original feature sets.

The results of feature selection using distance evaluation technique can be seen in Figs. 4.14 and 4.15. From this figures, we can see that there are 24 ICs and PCs are resulted from feature extraction process. Usually 5 to 12 parameters are sufficient to perform the calculation and provide sufficient accuracy [22]. Applying the distance evaluation technique remains 7 ICs and PCs which have largest distance evaluation criteria. The best ICs and PCs from feature selection are presented in Table 4.4.

Table 4.4 Selected ICs and PCs after feature selection

Independent components (ICs)	Principal components (PCs)
5, 10, 13, 14, 15, 18, 19	1, 2, 3, 4, 6, 13, 16

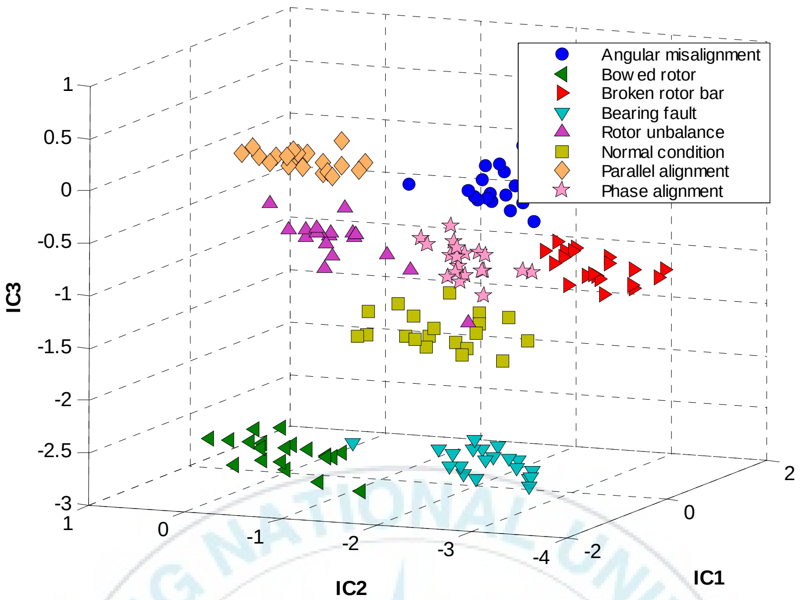


Fig. 4.12 Feature extraction using ICA.

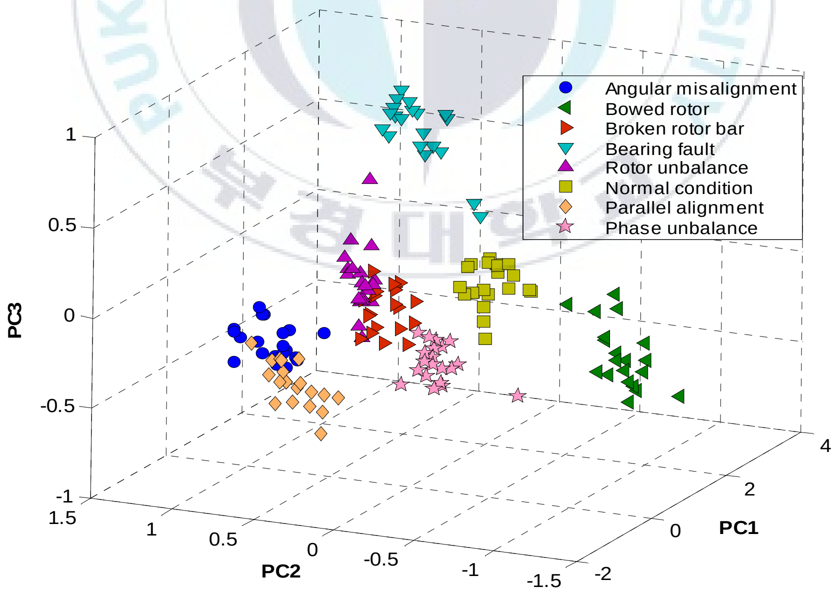


Fig. 4.13 Feature extraction using PCA.

5.1.5. Training and Classification

In this study, the RBF kernel and polynomial are used as the basic kernel function of SVM. 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 SVM. 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 [23].

The goal of this guideline is to identify optimal choice of C and γ so that the classifier can accurately classify the data input. In ν-fold cross-validation, we first divide the training set into subsets of equal size. Sequentially on subset is tested using the classifier trained on the remaining (ν-1) subsets. Thus, each instance of the whole of training set is predicted once so the cross validation accuracy is the percentage of data that are correctly classified. The cross-validation procedure can prevent the overfitting problem. In this dissertation, we use 10-fold cross validation to search the proper kernel parameter d, C, and γ. Basically, all the pairs of (C, γ) for RBF kernel and (d, C, γ) for polynomial kernel are tried and the one with the best cross-validation accuracy is selected. In this work, we performed the 10-fold cross-validation to choose the proper parameters of $C = \{2^0, 2^1, \ldots, 2^7\}$ and $\gamma = \{2^{-3}, 2^{-2}, \ldots, 2^3\}$.

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 Hsu and Lin [24]. The scenarios of training and classification process as follows: first, SVM based multi-class classification is trained on data input from original features without feature extraction and feature classification. Second, we change the data input for SVM training using data input after feature extraction by PCA and ICA. Furthermore, the variation of kernel function is performed to show the excellent of characteristic of kernel function and its performance in faults classification. In this work, we employed polynomial and Gaussian RBF kernel functions. Third, we retry the all of training and classification process by introducing kernel parameter selection. Finally, the results of the training and faults classification are compared to show the best results of the system.

5.1.6. Results and Discussion

The result of this study can be shown in Tables 4.5, 4.6, and 4.7. 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.

1. Effect of Feature Extraction and Selection

In Table 4.5, classification process is performed on the original feature set without feature extraction and selection. The classification rates of this process among 75.0% until 97.5%. The bad performance of this classification is due to the existence of irrelevant and useless features. Many irrelevant features make burden and tend to decrease the performance of classifier.

Then, as shown in Tables 4.6 and 4.7, the classification rate with PCA and ICA feature extraction ranged from 97.5% to 100%. It is better than the previous classification without feature extraction and selection. By using ICA and PCA feature extraction, the useful feature is extracted from original feature sets. Furthermore, the number of support vectors (SVs) decreased due to feature extraction. In this case, classification process using ICA feature extraction needs fewer numbers of SVs than PCA feature extraction and original feature. This phenomenon can be explained that ICA finds the components not merely uncorrelated but independent. Independent components are more useful for classification rather than uncorrelated components. The 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.

Moreover, from feature selection part, we can observe the effect of feature selection from the distance evaluation criteria of ICs and PCs. Fig. 4.14 show us that the variance of distance among the ICs is relatively high; it represents of useful ICs features. It means that the bigger variance of distance evaluation criteria have significant importance in classification process. From Fig. 4.15 we can see that the variance of distance between PCs is relatively low except first PCs. However, the others PCs have low variance in distance respectively. So, it indicates the performance of PCs is lower than ICs in classification process.

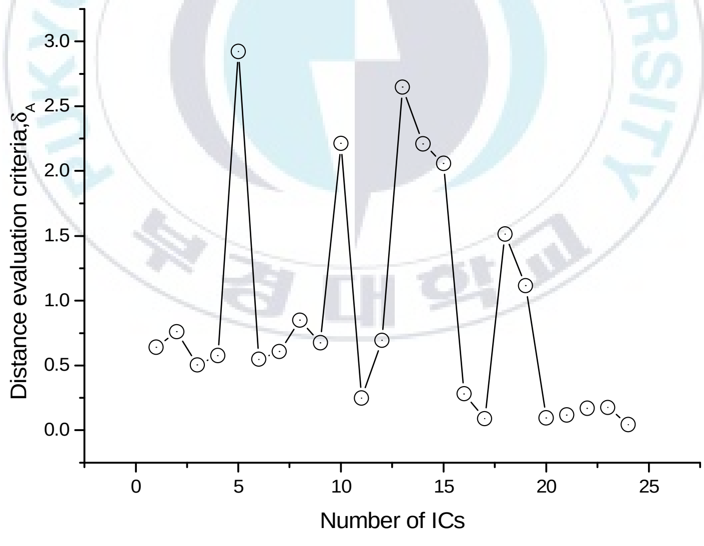


Fig. 4.14 Distance evaluation criteria of ICs.

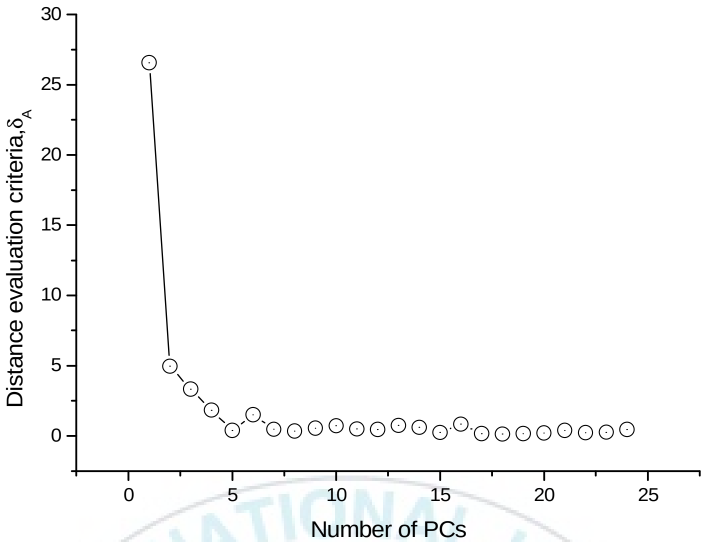


Fig. 4.15 Distance evaluation criteria of PCs.

Table 4.5 Fault classification using original feature and SVM

Kernel	Multi-class strategy	Classification rate (%) Training	Testing	Number of SVs	Training time (s)
Polynomial	One vs. one	89.2	90.0	93	0.48
(d = 1)	One vs. all	77.5	75.0	103	0.86
Polynomial	One vs. one	91.7	90.0	94	0.52
(d = 2)	One vs. all	81.7	80.0	95	0.56
Polynomial	One vs. one	93.3	97.5	93	0.56
(d = 3)	One vs. all	80.8	85.0	94	1.00
Polynomial	One vs. one	94.2	97.5	94	0.48
(d = 4)	One vs. all	80.0	98.5	94	0.98
Gaussian RBF	One vs. one	92.5	90.0	99	0.32
(γ= 2.19)	One vs. all	77.0	72.5	110	0.47

Table 4.6 Fault classification using PCA and SVM

Kernel	Multi-class strategy	Classification rate (%)		Number of SVs	Training time (s)
		Training	Testing		
Polynomial	One vs. one	100	100	79	0.52
(d = 1)	One vs. all	99.17	97.5	68	2.39
Polynomial	One vs. one	100	100	77	0.55
(d = 2)	One vs. all	100	100	84	2.17
Polynomial	One vs. one	100	100	72	0.48
(d = 3)	One vs. all	100	97.5	93	1.69
Polynomial	One vs. one	100	100	73	0.53
(d = 4)	One vs. all	100	97.5	96	2.37
Gaussian RBF	One vs. one	100	100	84	0.41
(γ= 2.19)	One vs. all	100	100	80	0.90

Table 4.7 Fault classification using ICA and SVM

Kernel	Multi-class strategy	Classification rate (%)		Number of SVs	Training time (s)
		Training	Testing		
Polynomial	One vs. one	100	100	48	0.31
(d = 1)	One vs. all	99.17	100	45	5.17
Polynomial	One vs. one	100	100	49	0.34
(d = 2)	One vs. all	100	97.5	45	0.92
Polynomial	One vs. one	100	100	45	0.32
(d = 3)	One vs. all	100	97.5	47	1.26
Polynomial	One vs. one	100	100	46	0.34
(d = 4)	One vs. all	100	97.5	52	1.15
Gaussian RBF	One vs. one	100	100	56	0.23
(γ= 2.19)	One vs. all	100	100	50	0.26

2. Effect of Kernel Function

From this study, the effect of selection of kernel function is also introduced. The performance of SVM depends on a great extent on the choice of kernel function to transform a data from input space to a higher dimensional feature space. The choice of kernel function is data dependent and there are no definite rules governing its choice that might yield a satisfactory performance. Tables 4.5, 4.6 and 4.7 present results of SVM with the kernel function defined in Table 3.1. In these tables, d is the degree of polynomial and γ is width of RBF kernel parameter. The parameter C does not emerge in this table because it only needed in calculation process as penalty term.

At the first classification, we do not optimize the kernel parameters. First, the polynomial kernel function was used and then the second we used Gaussian RBF kernel. RBF kernel is very popular and claimed as the best kernel in classification process. In this kernel, there are two parameters which determine the performance in training and testing, C and γ. Therefore, the selection of proper kernel parameters C and γ is very important to achieve the good performance. In this dissertation, we performed training and testing process using without or with kernel parameter selection. The effect of kernel parameters selection will be explained in the next discussion.

According to effect of kernel selection, the performance in classification training and testing tends to be increased using polynomial and RBF kernel, respectively. This phenomenon can be seen in the Tables 4.5, 4.6 and 4.7. The kernel parameters which used in polynomial kernel are d = 1, C = 10 and γ = 1. Whereas for RBF kernel we used C = 10 and γ = 2.19. In Table 4.5, the performance of RBF kernel using one-against-all strategy is lower than the others. This condition is caused by using improper RBF kernel parameters C and γ. Also, in Table 4.4, we used the original features without feature extraction and selection. That is why the performance of RBF kernel in Table 4.4 is lowest.

3. Effect of Kernel Parameters Selection

There are three parameters associated with polynomial kernel (d, C, γ) and two parameters for the RBF kernel (C, γ). It is not known beforehand which values of d, C and γ are the best for one problem; consequently, some kind of model selection or parameter search approach must be employed. This study conducts a 10-fold cross validation to find the best values of d, C and γ. Pairs of d, C and γ are tried and the one with lowest cross-validation error is picked. For RBF kernel we searched the range of parameters $C = \{2^0, 2^1, \ldots, 2^7\}$ and $\gamma = \{2^{-3}, 2^{-2}, \ldots, 2^3\}$, so there are 56 pairs of (C, γ) which must be evaluated. In the case of polynomial kernel we evaluated pairs of (d, C, γ) from the range $d = \{1, 2, 3, 4\}$, $C = \{2^0, 2^1, \ldots, 2^7\}$ and $\gamma = \{2^{-3}, 2^{-2}, \ldots, 2^3\}$. The polynomial kernel seems to have more hyper-parameters than RBF kernel. The complete results of kernel parameter selection are summarized in Table 4.8.

Table 4.8 Selected kernel parameter

Data	Polynomial kernel (d, C, γ)		RBF kernel (C, γ)	
	One vs. one	One vs. all	One vs. one	One vs. all
Original feature	$(3, 2^7, 2^0)$	$(4, 2^7, 2^0)$	$(2^7, 2^{-3})$	$(2^6, 2^{-3})$
PCA feature extraction	$(1, 2^2, 2^0)$	$(1, 2^0, 2^0)$	$(2^0, 2^{-1})$	$(2^0, 2^{-2})$
ICA feature extraction	$(1, 2^5, 2^0)$	$(1, 2^1, 2^0)$	$(2^1, 2^0)$	$(2^1, 2^0)$

After the optimal pairs were found, the whole training data was training again to generate the final classifier. This study performs the training process using polynomial and RBF kernel to all of data: original features, PCA feature extraction and ICA feature extraction. The performance of polynomial and RBF kernel after kernel parameter selection is presented in Tables 4.9, 4.10 and 4.11.

Table 4.9 Fault classification using original feature and selected kernel parameter

Kernel	Multi-class approach	Classification rate (%)		Number of SVs	Training time (s)
		Training	Testing		
Polynomial	One vs. one $(3, 2^7, 2^0)$	99.98	100	47	0.031
(d, C, γ)	One vs. all $(4, 2^7, 2^0)$	98.30	100	60	0.125
RBF	One vs. one $(2^7, 2^{-3})$	100	100	41	0.032
(C, γ)	One vs. all $(2^6, 2^{-3})$	100	100	43	0.078

Table 4.10 Fault classification using PCA and selected kernel parameter

Kernel	Multi-class approach	Classification rate (%)		Number of SVs	Training time (s)
		Training	Testing		
Polynomial	One vs. one $(1, 2^2, 2^0)$	100	100	47	0.031
(d, C, γ)	One vs. all $(1, 2^0, 2^0)$	100	100	91	0.063
RBF	One vs. one $(2^0, 2^{-1})$	100	99.97	71	0.016
(C, γ)	One vs. all $(2^0, 2^{-2})$	100	100	80	0.063

Table 4.11 Fault classification using ICA and selected kernel parameter

Kernel	Multi-class approach	Classification rate (%)		Number of SVs	Training time (s)
		Training	Testing		
Polynomial	One vs. one $(1, 2^5, 2^0)$	100	100	42	0.031
(d, C, γ)	One vs. all $(1, 2^1, 2^0)$	100	100	79	0.062
RBF	One vs. one $(2^1, 2^0)$	100	100	64	0.015
(C, γ)	One vs. all $(2^1, 2^0)$	100	100	79	0.063

As shown in Tables 4.9–4.11, the performance of classification process is increased due to the kernel parameter selection. It can be compared with Tables 4.5–4.7 in the case of without kernel parameter selection. In Table 4.9, the classification rates of training for polynomial kernel is lower than RBF kernel both one-against-one and one-against-all strategies even though the degree of polynomial are 3 and 4 respectively. This condition is supposed due to the bad quality of data input without feature extraction so that the curse of dimensionality phenomenon decreases the performance of classifier. However, in Tables 4.10 and 4.11, the classification rate reaches 100% using polynomial kernel due to good quality of data input after feature extraction process.

In Table 4.10, the classification rates of each kernel function are high; even almost of classification rates achieve 100%. Generally, the strategy of one-against-one is better than one-against-all as listed in the table. As shown in Table 4.10, the feature extraction using PCA is useful to increase the performance of classification rather than without feature extraction in Table 4.9, because of PCA search the uncorrelated components from the input data space and treat it so that more useful in classification. Moreover, using kernel parameter selection will increase the performance better. The proper pairs of (d, C, γ) in polynomial kernel are (1, 2^2, 2^0) and (1, 2^0, 2^0) for one-against-one and one-against-all respectively. Although the degree of polynomial equal to 1, however, the performance is high (100%). In RBF kernel, the proper kernel parameter of pairs (C, γ) are (2^0, 2^{-1}) and (2^0, 2^{-2}) for one-against-one and one-against-all respectively. The classification rates also high (100% and 99.97%). It becomes evidence that kernel parameter selection is very important to get good performance. Furthermore, the use of proper kernel parameter will overcome the problems of underfitting and overfitting so the best classification process is yielded.

Finally, the faults classification using ICA feature extraction is presented in Table 4.11. This table presents the best performance in faults classification rather than previous methods. From this table we can see that performance of all kernel function are 100% in fault classification. It shows us that the feature extraction using ICA is the best method among them, because of ICA seeks not merely uncorrelated components but independents. It is more useful for classification process. In addition, the application of kernel parameter selection using cross-validation makes the performance of classification is excellent. The results of kernel parameter selection for polynomial kernel (pairs of (d, C, γ)) are (1, 2^5, 2^0) and (1, 2^1, 2^0) for one-against-one and one-against-all respectively. It uses one degree of polynomial kernel and produces the best performance. Then, in the case of RBF kernel, the kernel parameter selection pairs of (C, γ) yields (2^1, 2^0) both for one-against-one and one-against-all respectively. The classification rate is more excellent, both 100% rather than PCA feature extraction and original feature.

5.2. Case Study 2: Using Nonlinear Feature Extraction

In this case study, experiment is conducted using same test rig and data acquisition method to collect the data. The features used in this case study are generated using same feature calculation method. A nonlinear feature extraction method is performed by introducing kernel function in previous linear method of feature extraction.

5.2.1. Feature Extraction

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. 4.16.

Fig. 4.16 plots 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 should not be directly processed into classifier because it will degrade the performance of classifier.

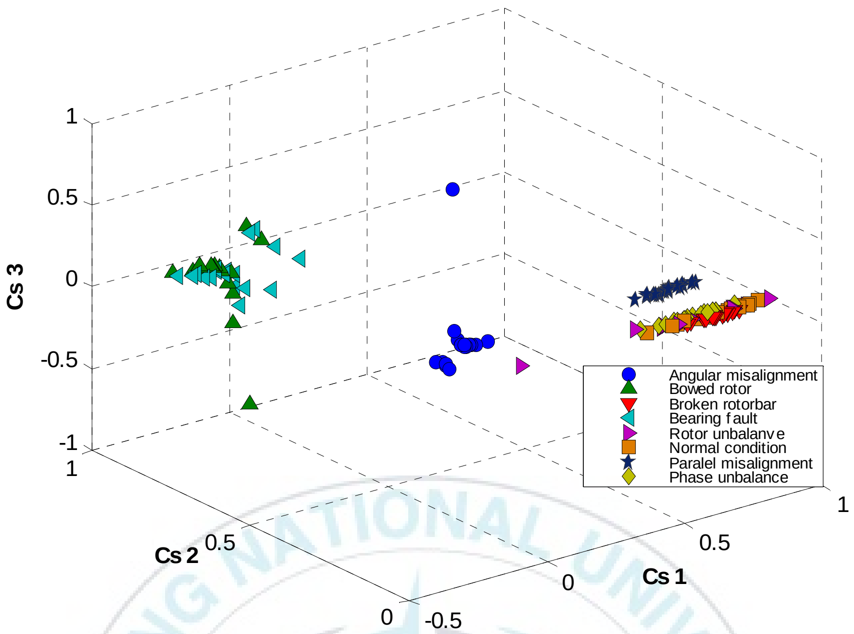


Fig. 4.16 Three-first components of original data features.

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 the disorder structure of data features. In this work, the use of kernel PCA for feature extraction is introduced. Based on the eigenvalue, we select 97% of the total largest eigenvalue of covariance matrix as a reference to reduce the dimensionality. Representation of eigenvalue can be seen in Fig. 4.17 which presents 20 largest eigenvalues of covariance matrix. Then, we select the RBF kernel function in kernel PCA and choose the kernel parameter γ= 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 presented in Figs. 4.18 and 4.19.

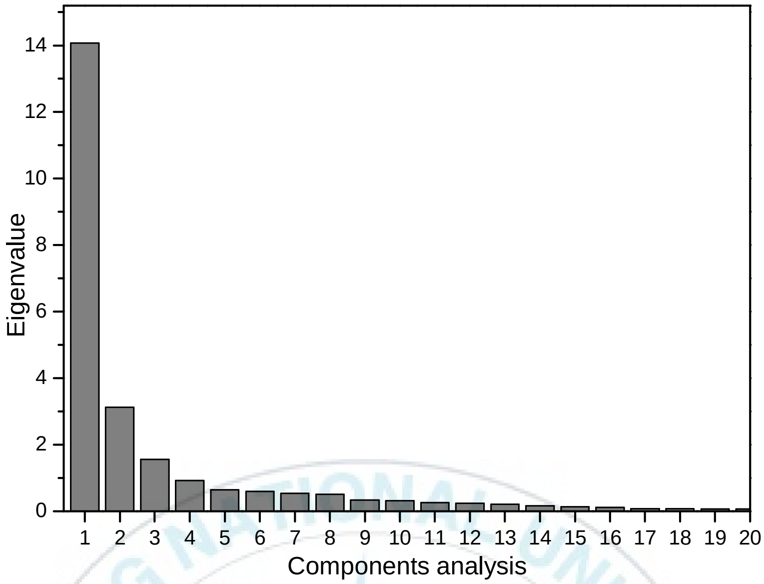


Fig. 4.17 Representation of 20 eigenvalues of covariance matrix.

In Fig. 4.18, we can see that kernel PCA 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 kernel PCA in clustering is associated that kernel PCA can explore higher order information of the original data feature beside of uncorrelated data.

In the next step, we performed nonlinear feature extraction using whitened data feature by kernel PCA and employed ICA algorithm to seek the projection direction in kernel PCA whitened space. We called this process as kernel ICA feature extraction. In this technique, we expect that feature extraction process can be improved due the robustness of kernel ICA.

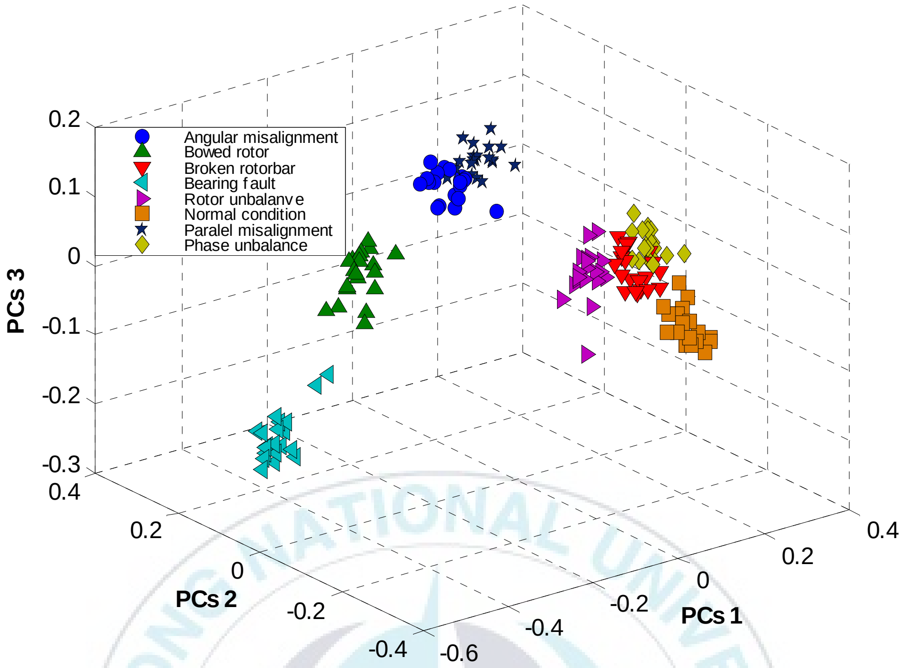


Fig. 4.18 Feature extraction using kernel PCA.

Fig. 4.19 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 kernel ICA is the best in comparing with previous technique. In addition, kernel ICA also implicitly takes into account the high order information of the original data features. Furthermore, in kernel ICA technique, the mutual independent components will give the promising to be a useful and the best features.

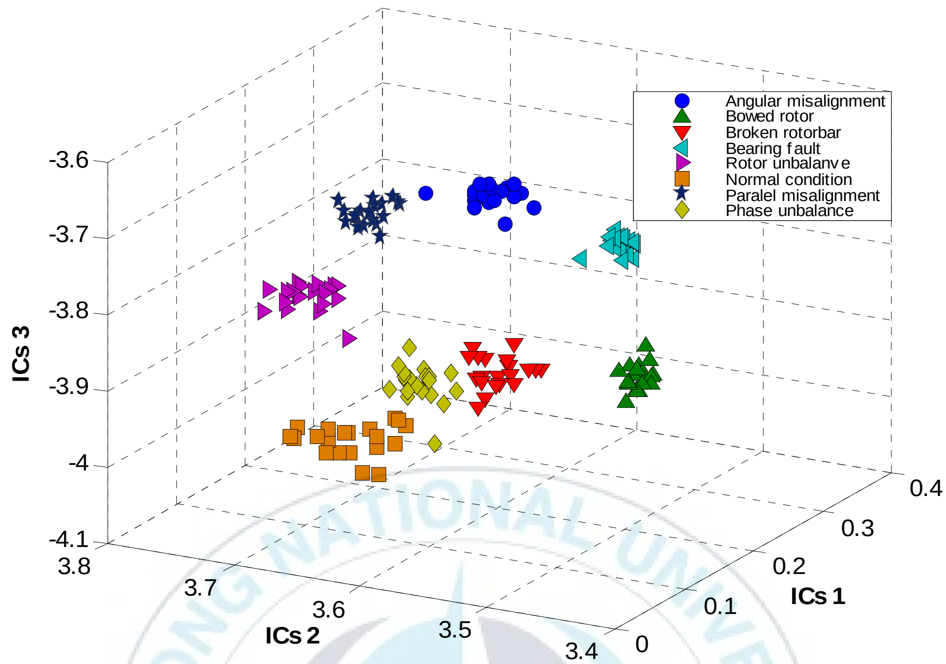


Fig. 4.19 Feature extraction using kernel ICA.

To investigate the performance of nonlinear feature extraction process using kernel PCA and kernel ICA, we calculated the average of Euclidean distance between points in class of feature space [25,26]. 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 kernel PCA and kernel ICA feature space respectively then took the lowest which represents the good clustering. The calculation of average Euclidean distance can be seen in Fig. 4.20. In this figure we can see that the average distance of kernel ICA is lower than kernel PCA so it becomes evidence that performance of kernel ICA significantly outperforms kernel PCA in terms of clustering.

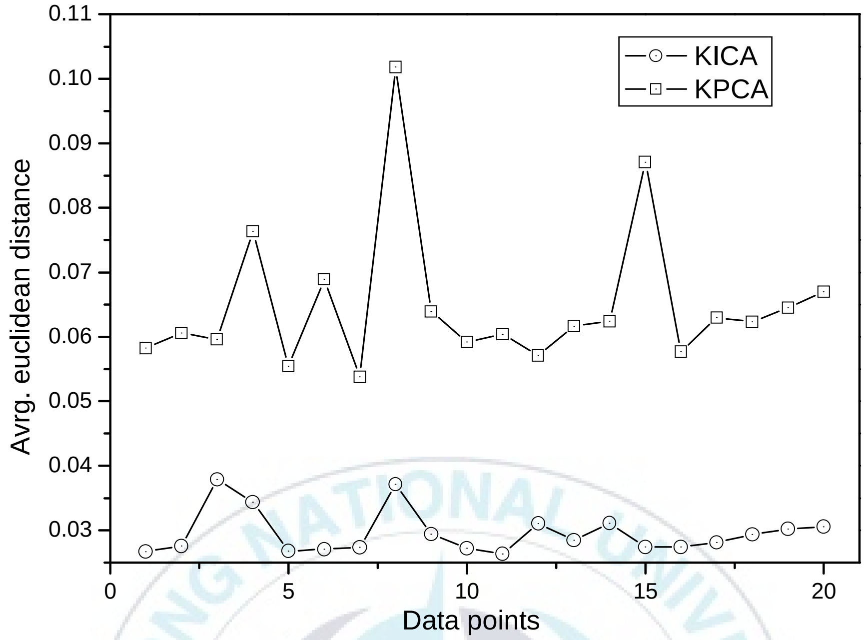


Fig. 4.20 Average of Euclidean distance kernel PCA and kernel ICA.

5.2.2. 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 [24]. To solve the SVM problem, Vapnik [27] describe a method which used the projected conjugate gradient algorithm to solve the SVM-QP problem. Sequential minimal optimization (SMO) proposed by Platt [28] 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 dissertation, SMO is used as a solver.

In this study, we use 10-fold cross-validation to search the proper kernel parameter d, C, and γ. Basically, all the pairs of (C, γ) for RBF kernel and (d, C, γ) for polynomial kernel are tried and the one with the best cross-validation accuracy is selected. We perform the 10-fold cross-validation to select the proper parameters of $C = \{2^0, 2^1, \ldots, 2^7\}$ and $\gamma = \{2^{-3}, 2^{-2}, \ldots, 2^3\}$.

There are three parameters associated with polynomial kernel (d, C, γ) and two parameters for the RBF kernel (C, γ). It is not known beforehand which values of d, C and γ are the best for one problem; consequently, some kind of model selection or parameter search approach must be employed. This study conducts a 10-fold cross validation to find the best values of d, C and γ. Pairs of d, C and γ are tried and the one with lowest cross-validation error is picked. For RBF kernel we searched the range of parameters $C = \{2^0, 2^1, \ldots, 2^7\}$ and $\gamma = \{2^{-3}, 2^{-2}, \ldots, 2^3\}$, so there are 56 pairs of (C, γ) which must be evaluated. In the case of polynomial kernel we evaluated pairs of (d, C, γ) from the range $d=\{1, 2, 3, 4\}$, $C = \{2^0, 2^1, \ldots, 2^7\}$ and $\gamma = \{2^{-3}, 2^{-2}, \ldots, 2^3\}$. The polynomial kernel seems to have more hyper-parameters than RBF kernel. The complete results of kernel parameter selection are summarized in Table 4.12.

Table 4.12 Selected kernel parameter

Feature extraction method	Polynomial kernel (d, C, γ)		RBF kernel (C, γ)	
	One vs. one	One vs. all	One vs. one	One vs. all
Kernel PCA	$(3, 2^2, 1)$	$(3, 2^4, 1)$	$(2^7, 2^{-1})$	$(2^7, 2^{-2})$
Kernel ICA	$(1, 2^7, 1)$	$(1, 2^6, 1)$	$(2^7, 2^{-2})$	$(2^6, 2^{-1})$

5.2.3. Results and Discussion

Table 4.13 presents the result of classification using kernel PCA feature extraction and SVM. According to the accuracy, this method is very good because all of classification accuracies are 100%. The excellent of this method is also shown by the number of SVs which is reduced and smaller than previous method except the one-against-one strategy of polynomial kernel. However, according to training time, the classification process using kernel PCA feature extraction and

SVM is relatively longer than linear feature extraction. In addition, this table shows that the strategy of one-against-one is better than one-against-all strategy as previous methods.

The proper kernel parameters are $(3, 2^4, 1)$ and $(3, 2^2, 1)$ for polynomial kernel of one-against-all and one-against-one strategy respectively. In this case, it means that the feature extraction method using kernel PCA needs high degree of polynomial (d = 3) for classification process of SVM. And it becomes a reason that the training time is longer than PCA and ICA feature extraction with same kernel. The proper parameters of RBF kernel are $(2^6, 2^{-1})$ and $(2^7, 2^{-2})$ for one-against-all and one-against-one strategy, respectively. Classification process using kernel PCA feature extraction and RBF kernel has reduced the number of SVs smaller than previous method.

The result of classification using kernel ICA feature extraction and SVM is presented in Table 4.14. The accuracy of this process also high which reached 100% except one-against-all strategy using polynomial kernel 99.97%. Generally, in comparing with feature extraction using kernel PCA the performance of this method is better according to the number of SVs and training time. Application of kernel parameter selection using cross-validation method produced the proper kernel parameters which made its performance is excellent. For polynomial kernel, the proper parameters are $(1, 2^6, 1)$ and $(1, 2^7, 1)$ for one-against-all and one-against-one strategy respectively. Although the degree of polynomial kernel equal to 1 however it reached good performance. The most excellent of performance is shown using RBF kernel which produced high accuracy and smallest number of SVs. This excellence is surely influenced by the use of proper kernel parameters that are $(2^6, 2^{-1})$ and $(2^7, 2^{-2})$ for one-against-all and one-against-one strategy respectively.

Table 4.13 Fault classification using kernel PCA and SVM

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

Table 4.14 Fault classification using kernel ICA and SVM

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

5.3. Case Study 3: Motor Current Signal and W-SVM

This method is well known as motor current signal analysis (MCSA) which use stator current signal of motor induction to conduct fault diagnosis. A brief review discussing how to use MCSA was highlighted in [3,29,30]. In present study, the start-up transient current signature is selected for detection and diagnosing of faults in induction motor. This method is effective because the machine is subjected to more stresses during the start-up above those of normal operation. These stresses can highlight the machine defects those are early in their development and not detected easily during steady state operation. The other advantage is that the transient signal has a high slip and high signal-to-noise ratio (SNR), which implies that its spectral can be detected more easily. Therefore, it is no need to heavily load the motor in order to make an accurate fault diagnosis.

5.3.1. Experiment and Data Acquisition

The experiment was conducted using test rig described in section 6.1.1. However, we only used three current probes to acquire the transient current signal. The maximum frequency of the used signals was 5 kHz. The data was acquired using data acquisition unit of 16 bit resolution; the number of the sampled data was 16384 at a sampling rate of 12800 Hz. The high sampling rate is selected because the transient signal has short time duration while sufficient samples per second are needed. Moreover it is planned to investigate and take the features from high frequency range after preprocessing using wavelet transform. One of the motors is normal condition (healthy) to be used as a benchmark for comparing with faulty condition. The conditions of faulty motors are described in Table 4.2.

5.3.2. Signal Preparation and Feature Calculation

The signals can be divided into two types; stationary and non-stationary. Transient signal which starts and finishes at zero is categorized as non-stationary signal. In this study, the start-up signal of induction motor is considered as transient signal (Fig. 4.21).

Working on the statistical features, smoothing process is necessary in order to remove or reduce the line frequency to highlight the differences of faults in induction motors. After smoothing, the transient start-up signal is expected to be similar to sine waveform which has variable amplitude. A moving average filter is used to smooth the data by replacing each data point with the average of the neighboring data points defined within a window. This window moves across the data set as the smoothed response value is calculated for each predictor value [31]. Subtracting the smoothed signal from the original signal gives the residual part which contains the information related to the normal or faulty conditions in induction motor. This residual part of transient current is shown in Fig. 4.22.

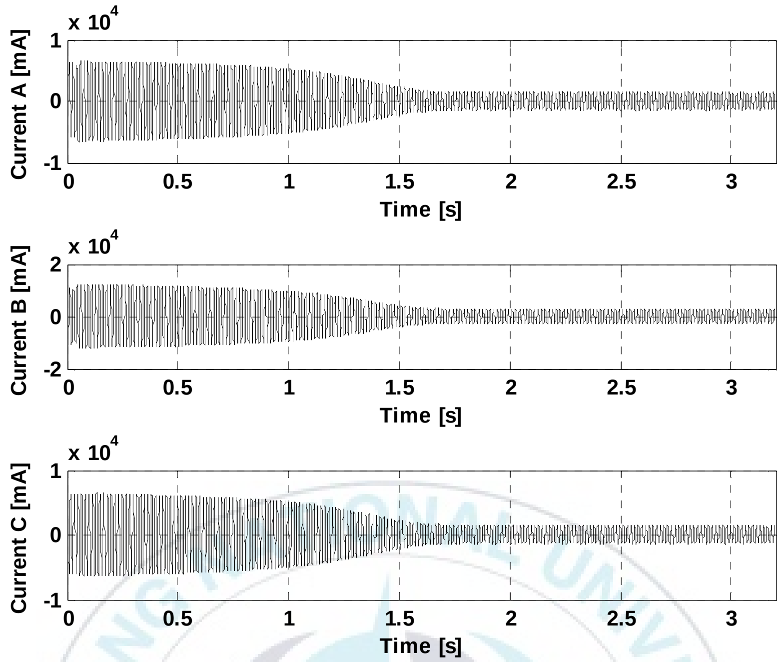


Fig. 4.21 Transient start-up current of phase A, B and C.

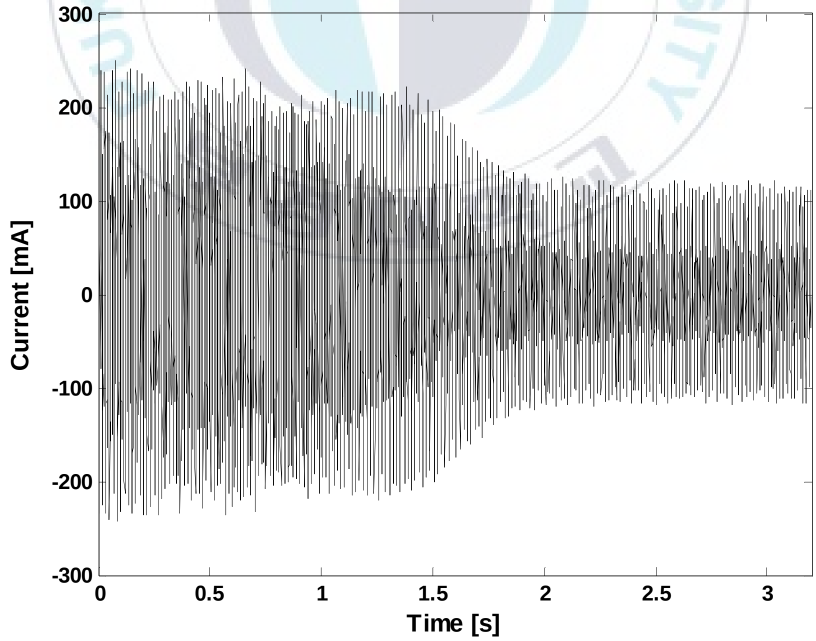
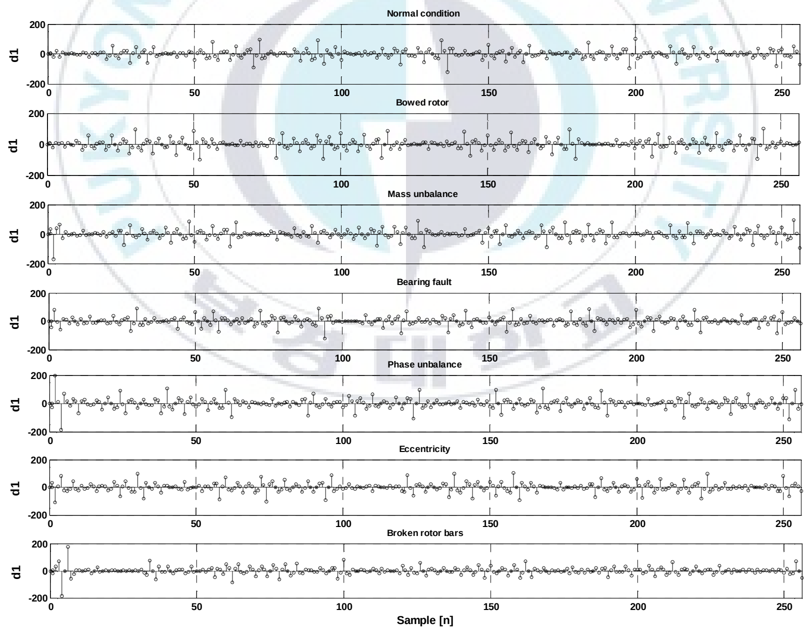


Fig. 4.22 The residual part of transient current.

Next step, discrete wavelet transform (DWT) is performed to extract the differences of each condition of induction motor. We performed 5 decomposition levels using Daubechies 5 (Db5) to show the salient features of faults in some frequency ranges. Wavelet transform can be considered as band pass filter where the different levels corresponds to the frequency at which different fault can be highlighted. The results of discrete wavelet transform are shown in Fig. 4.23.

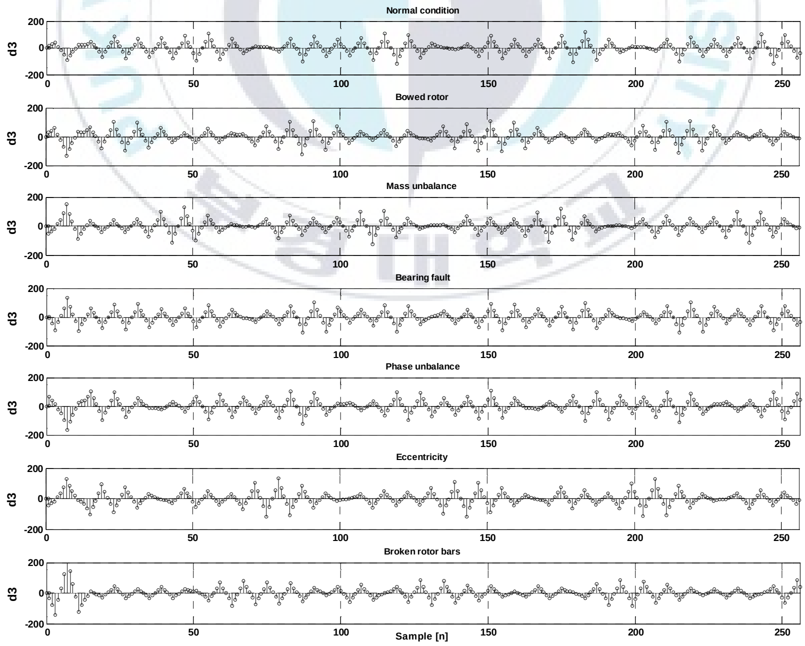
Fig. 4.23 shows the five decomposition levels of each condition of induction motor using DWT. Even though the differences between each condition are not clear, we can select one of them for feature calculation and left the others for reducing the dimensionality. In this dissertation, level 1 (d1) contained in the high frequency range is selected as the features source for classification process.



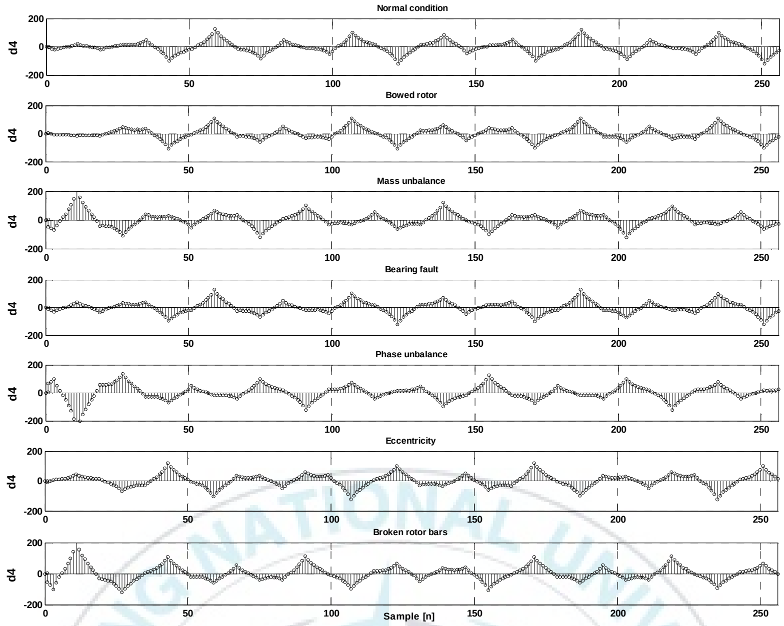
(a)


Normal condition
Bowed rotor
Mass unbalance
Bearing fault
Phase unbalance
Eccentricity
Broken rotor bars
d2
Sample [n]

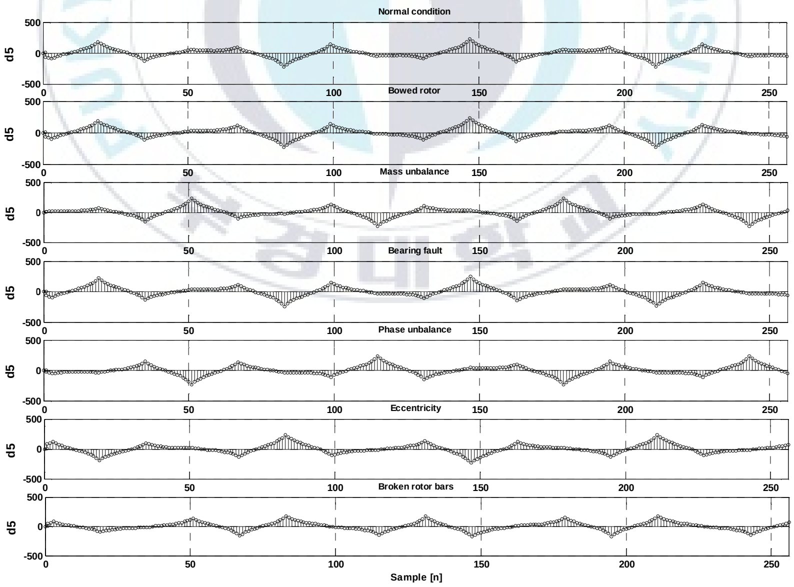
(b)


Normal condition
Bowed rotor
Mass unbalance
Bearing fault
Phase unbalance
Eccentricity
Broken rotor bars
d3
Sample [n]

(c)



(d)



(e)

Fig. 4.23 Wavelet transform for transient start-up signal of induction motor: (a) d1, (b) d2, (c) d3, (d) d4 and (e) d5.

After wavelet transform, the features are calculated from time waveform, frequency domain and auto regression of d1. In total, 63 features are obtained from 21 feature parameters of the 3 phases A, B, and C. A total of 140 data calculated from 7 conditions and each one has 20 measurements. The detailed of features are listed in Table 4.2.

5.3.3. Feature Extraction

In this study, non-linear feature extraction using kernel function is proposed to obtain good features for classification process. After feature calculation, the mean, rms and shape factor are plotted in Fig. 4.24 in order to know the structure of the data features. Fig. 4.24 can be a representation of the original features which have disorder structure or overlap and are not well clustered. Plotting original feature parameters indicates the necessity of preprocessing of the original features to make them better and ready for classification.

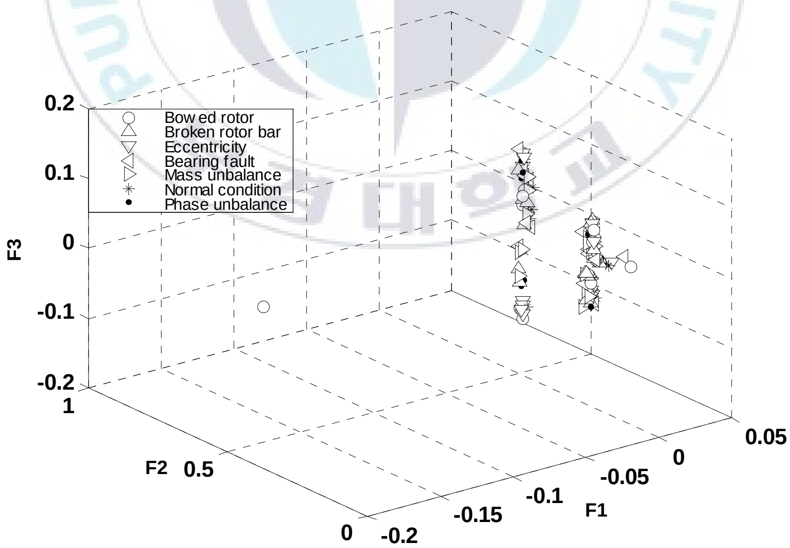


Fig 4.24 Original features.

Disorder structure of original features tends to decrease the performance of classifier if it is directly processed in classifier. To avoid this disadvantage, component analysis using PCA and KPCA are used to extract and reduce the feature dimensionality based on eigenvalue of covariance matrix. Fig. 4.25 shows the feature reduction in PCA and KPCA based on eigenvalue of covariance matrix. The features are changed into principal components and remaining only five for classification process.

The principal components of PCA and KPCA are plotted in Fig. 4.26. It can be observed that the clusters for seven conditions are not separated well. There are still overlapping among each condition of motor. It indicates that the features which are produced by current signature are very difficult to cluster. Therefore, more advance and good preprocessing is needed so that the salient differences features can be explored and emerged.

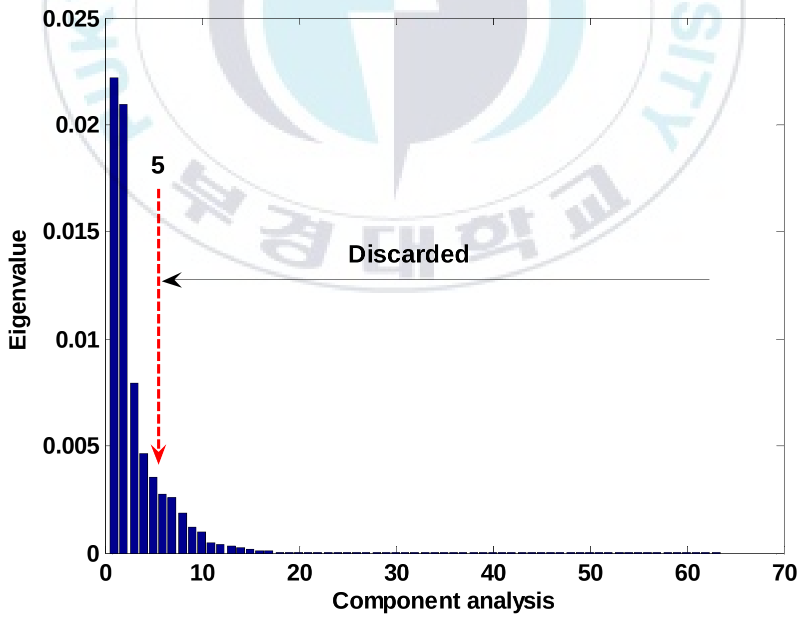
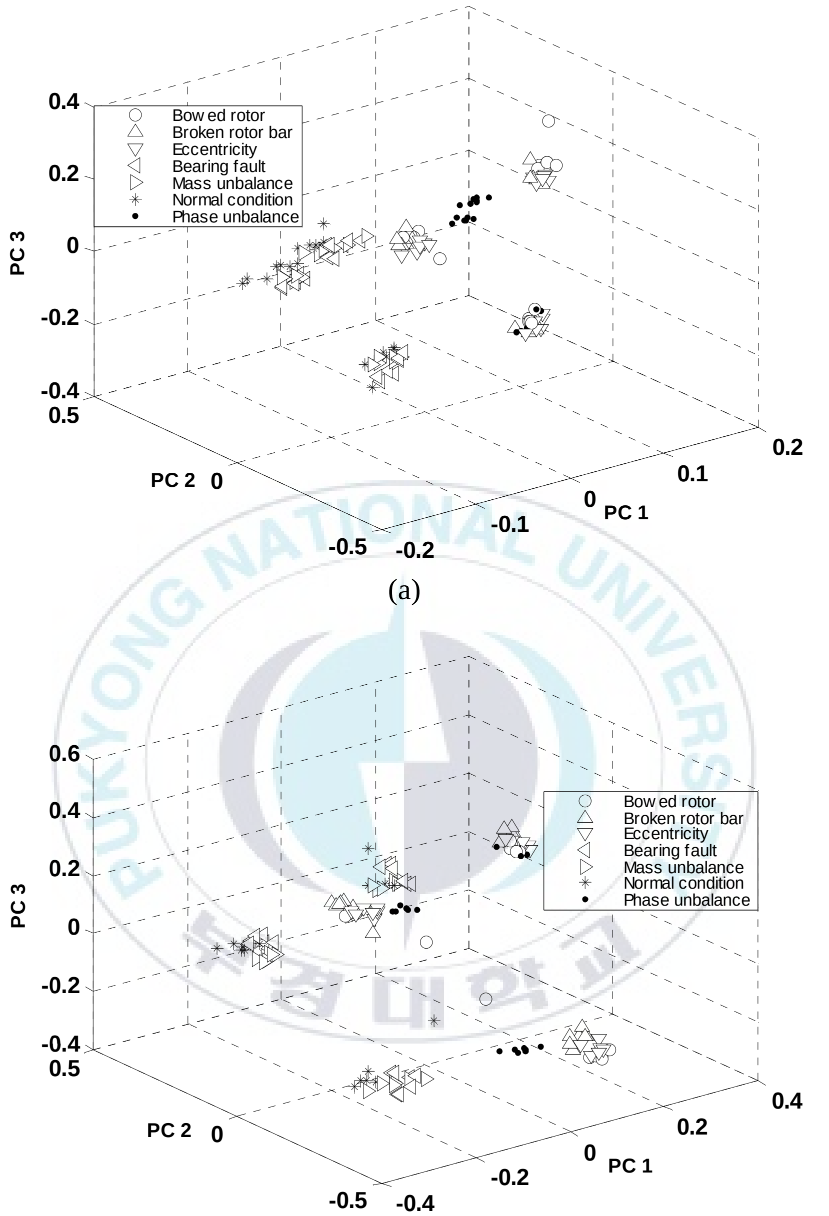


Fig. 4.25 Eigenvalue of covariance matrix for feature reduction.



(b)

Fig. 4.26 Principal components and kernel principal components.

5.3.4. Training and Classification

The SVM based on multi-class classification is applied to perform the classification process using one-against-all methods. The tutorial of this method was clearly explained in [24]. To solve the SVM problem, Vapnik described a method which used the projected conjugate gradient algorithm to solve the SVM-QP problem [28]. In the present study, SVM-QP is performed to solve the classification problem of SVM. The parameter C (bound of the Lagrange multiplier) and λ (condition parameter for QP method) are assigned the values 1 and 10^{-7}, respectively. Wavelet kernel function using Haar, Daubechies and Symlet are performed in the present study. The parameter δ in wavelet kernel refers to number of vanishing moment and is set to 4.

5.3.5. Results and Discussion

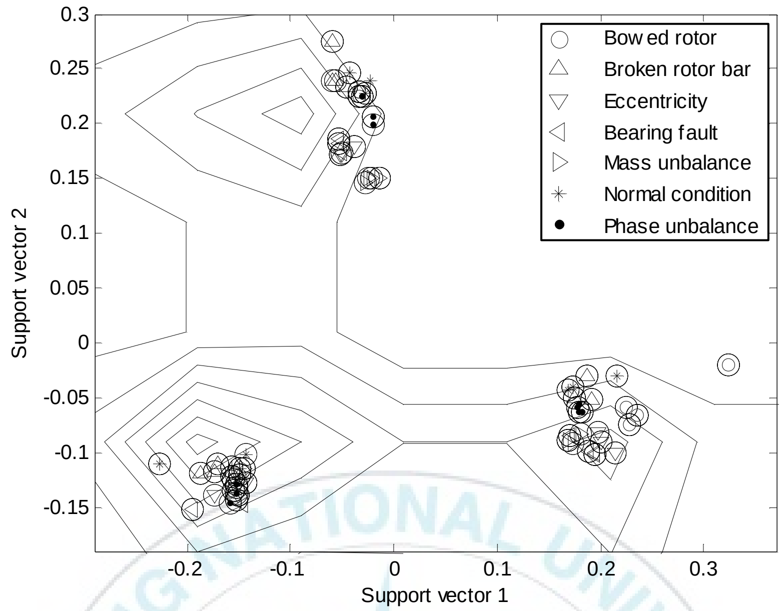
The complex boundaries of separation are presented in Figs. 4.27-4.29, from which the separation of W–SVM can be shown. In these figures, the circle refers to the support vector that states the correct recognition in W–SVM. Each condition of induction motor is well recognized using wavelet kernel except Haar kernel of PCA. Although W–SVM is well performed in recognition, however, each condition cannot be clustered and separated well. This phenomenon appears in all wavelet kernel functions. The lack of performance in preprocessing the transient current signal is suspected to be a reason why each class cannot be clustered well. Furthermore, because of the difficulty of handling the start-up transient current signal, so it needs an advanced preprocessing method. In this dissertation, the use of moving averaging and DWT (Db5) for preprocessing is not sufficient to emerge the salient differences between conditions in induction motors. Hence a proper preprocessing for the transient current signal is needed to be further investigated. Even though the clustering is not performed well, however, the correct classification and recognition show good performance using W–SVM. It is evident that W–SVM performs well in faults detection and classification of induction motors.

The performance of classification process summarized in Table 4.15 uses conventional kernel functions such as Gaussian and Polynomial for comparison purpose. All the data sets come from component analysis are accurately classified using wavelet kernel function, except Haar wavelet which reveals accuracies of 85% and 95% for PCA and KPCA, in training and testing accuracy, respectively. Wavelet kernel using Daubechies and Symlet reach accuracy 100% in training and testing, respectively. The number of support vectors are 68, which is relatively high due to the less quality of input data. The CPU time of Daubechies and Symlet wavelet are 7.1 s, 10.9 s and 9.6 s, 15.8 s for PCA and KPCA, respectively. These are higher than the CPU time of Haar wavelet which amounts to 5.8 s for both PCA and KPCA.

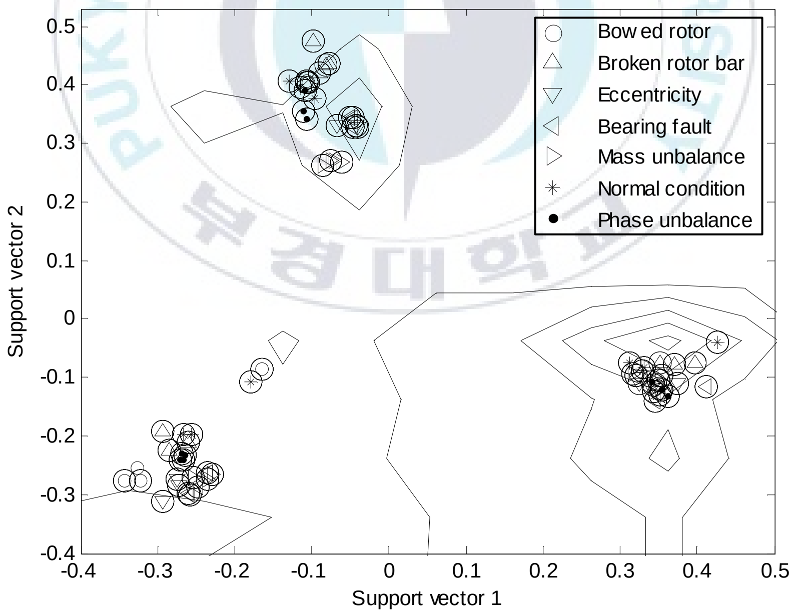
Table 4.15 Classification results

		W–SVM			Conventional SVM	
		Haar	Daubechies	Symlet	Gaussian (γ= 0.25)	Polynomial (d = 3)
Accuracy (%) (training/test)	PCA	85/85	100/100	100/100	75/75	61/61
	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

Table 4.15 shows the performance of conventional kernel function such as Gaussian and Polynomial. The accuracies are lower comparing with wavelet kernel function for all component analysis methods using PCA and KPCA. The number of support vectors is 70 higher than wavelet kernel function. However, the time consumptions are less than wavelet kernel function.

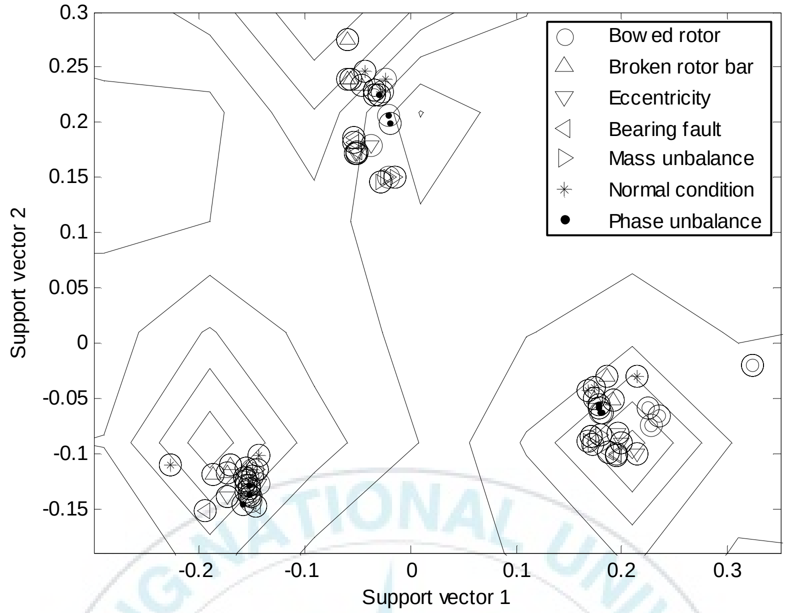


(a)

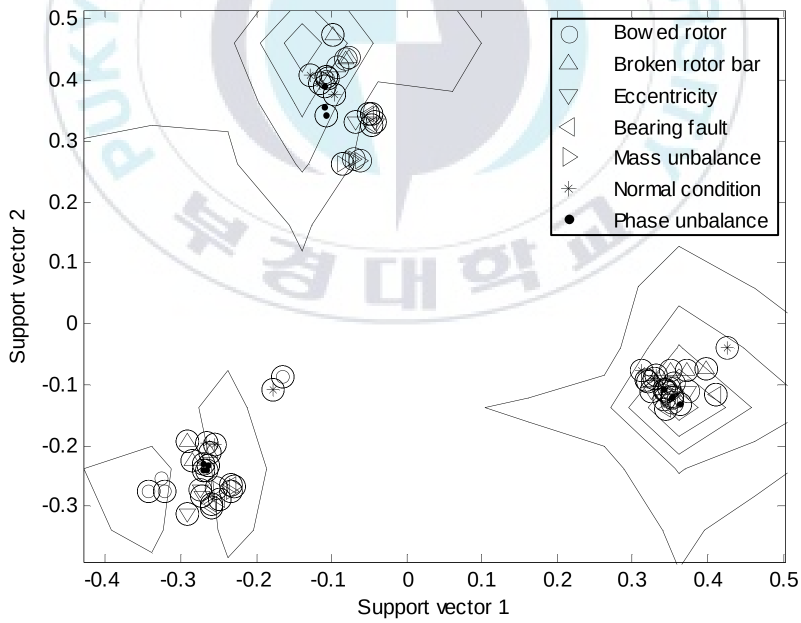


(b)

Fig. 4.27 Boundaries of separation using Haar wavelet kernel: a) PCA, b) KPCA.

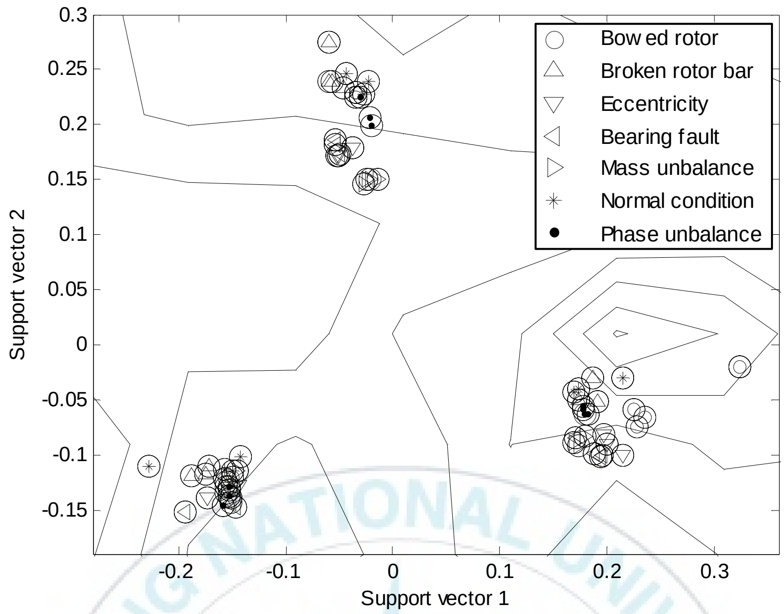


(a)

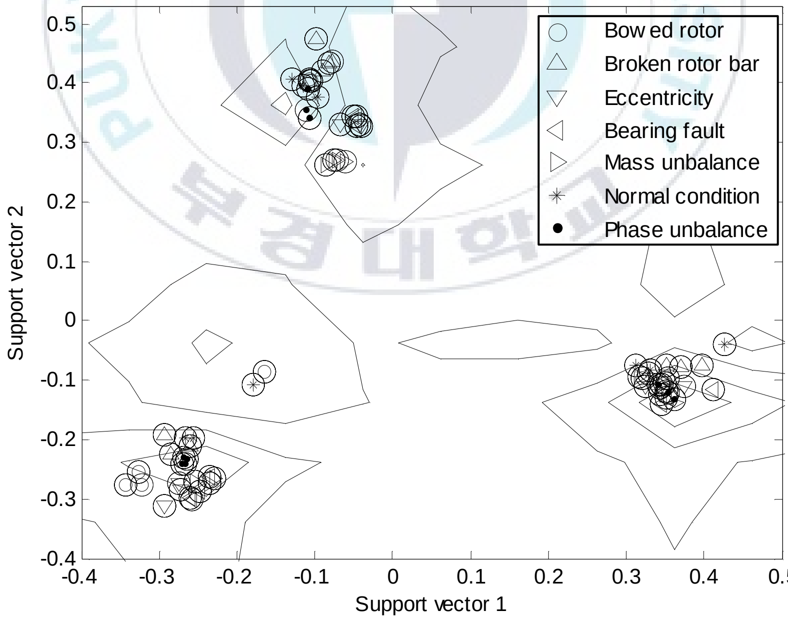


(b)

Fig. 4.28 Boundaries of separation using Daubechies wavelet kernel: a) PCA, b) KPCA.



(a)



(b)

Fig. 4.29 Boundaries of separation using Symlet wavelet kernel: a) PCA, b) KPCA.

In this study, as conclusion, a relatively new method of nonlinear kernel based on wavelet (W−SVM) has been introduced. The kernel function transforms the data into higher dimensional space in order to make it possible to perform the separation process. Feature reduction and extraction using component analysis via PCA and KPCA are highlighted. The performance of W−SVM is validated by applying it to faults detection and classification of induction motor based on start-up transient current signal. The results show that W−SVM is well performed and reached high accuracy in training and testing process based on experimental work. However, a proper preprocessing for the transient current signal is needed to improve emerging the salient differences between conditions in induction motors. Introducing nonlinear kernel using wavelets is believed to improve significantly the SVM research fields.

5.4. Case Study 4: Vibration Signal and W-SVM

5.4.1. Experiment and Data Acquisition

Data acquisition was conducted on induction motor of 160 kW, 440 volt, 2 poles as shown in Fig. 4.30. Six accelerometers were used to pickup vibration signal at drive-end and non drive-end on vertical, horizontal and axial direction, respectively. The maximum frequency of the used signals and the number of sampled data were 60 Hz and 16384, respectively.

5.4.2. Feature Calculation

The condition of induction motor is briefly summarized in Table 4.16. Each condition was labeled as class from 1 to 7. Feature representation for training and classification was presented in Table 4.3. There are totally 126 features calculated from 6 signals, 21 features and 98 data calculated from 7 condition 14 measurements.

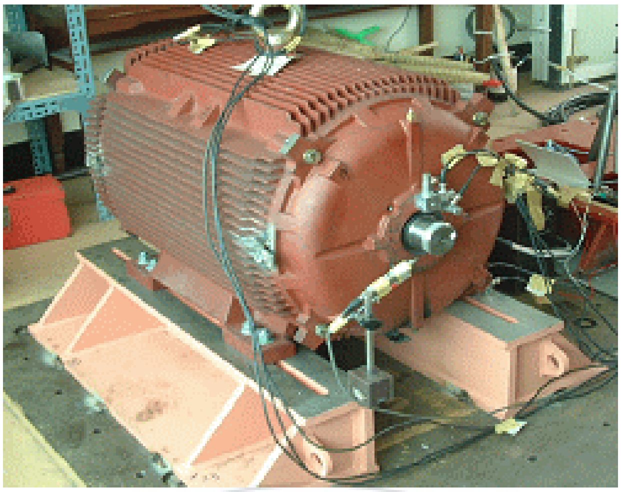


Fig. 4.30 Data acquisition of induction motor.

Table 4.16 Condition of induction motor

Class No.	Condition	Description	Others
1	Bent rotor	Maximum shaft deflection	1.45mm
2	Eccentricity	Static eccentricity (30%)	Air-gap: 0.25 mm
3	MCDE	Magnetic center moved (DE)	6 mm
4	MCNDE	Magnetic center moved (NDE)	6 mm
5	Normal	No faults	-
6	Unbalance	Unbalance mass on the rotor	10 gr
7	Weak-end shield	Stiffness of the end-cover	-

5.4.3. Feature Extraction and Reduction

Basically feature extraction is mapping process of data from higher dimension into low dimension space. This step is intended to avoid the curse dimensionality phenomenon. Structure of three first original features, those are mean, RMS and shape factor are plotted in Fig. 4.31. This figure shows the performance of original features those are containing overlap in some conditions. Then, applying component analysis is suggested to make original features well clustered.

Component analysis via ICA, PCA and their kernel are then used to extract and reduce the feature dimensionality based on eigenvalue of covariance matrix as described in Fig. 4.32. After performing component analysis the feature have been changed into independent and principal components, respectively. The first three independent and principal components from PCA, ICA and their kernel are plotted in Fig. 4.33. It can be observed that the clusters for seven conditions are separated well. It indicates that component analysis can perform feature extraction and all at once do clustering each condition of induction motors.

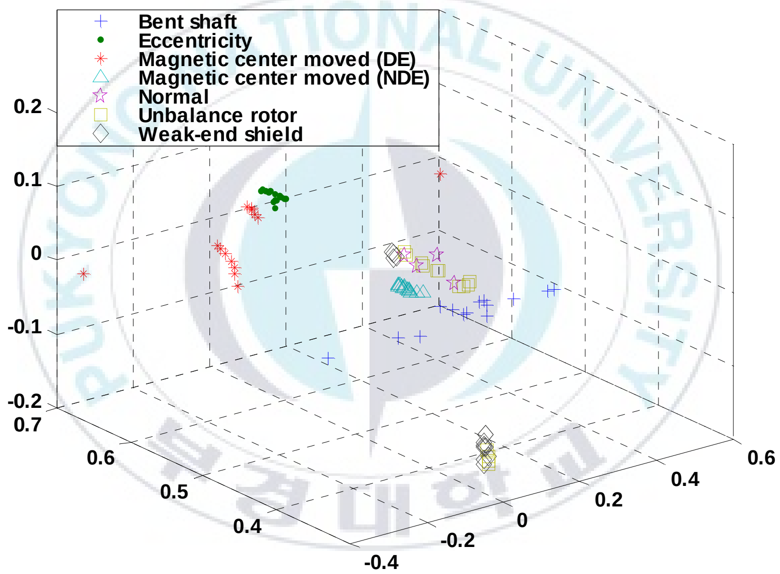


Fig. 4.31 Original features.

According to the eigenvalue of covariance matrix, the features were changed into component analysis and reduced only 5 component analysis needed for classification process. The other features are discarded due to small of eigenvalue of covariance matrix. The selected component analysis is then used by W-SVM classifier as input vectors for fault diagnosis using classification routine.

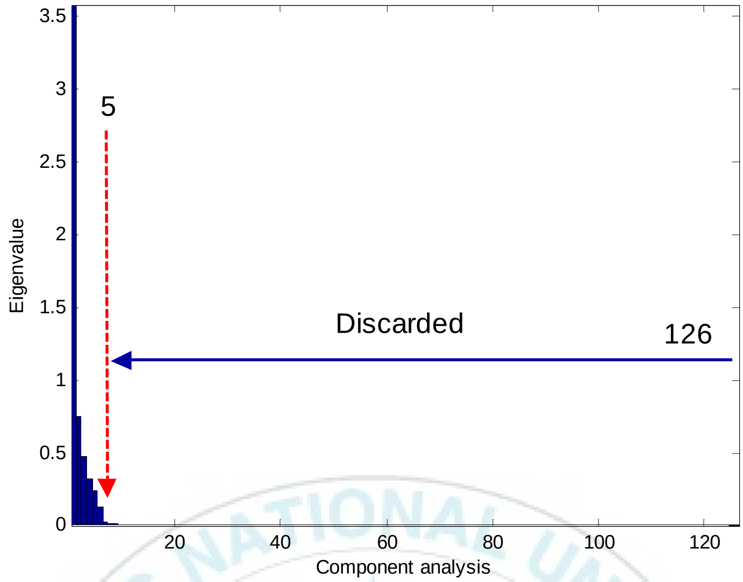
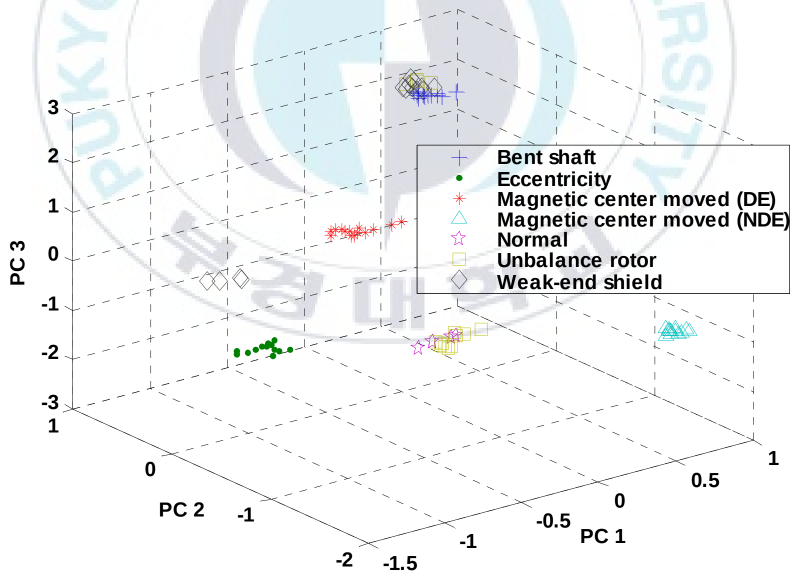
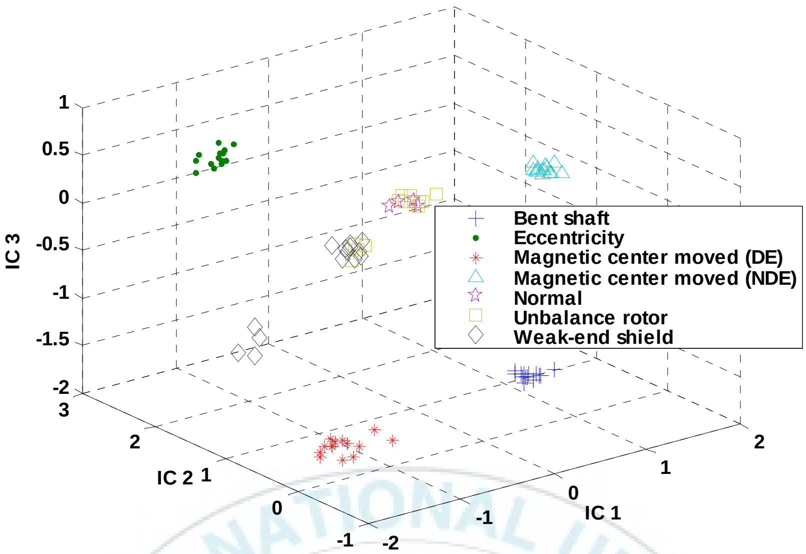


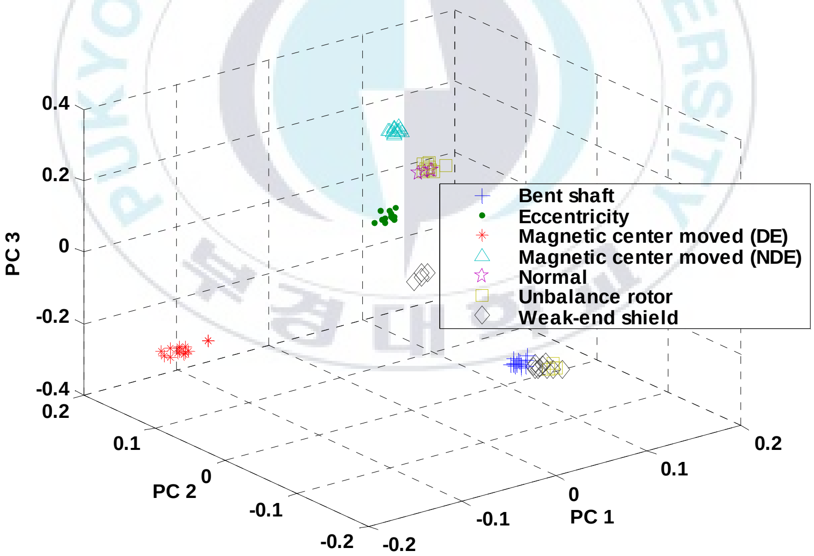
Fig. 4.32 Feature reduction using component analysis.



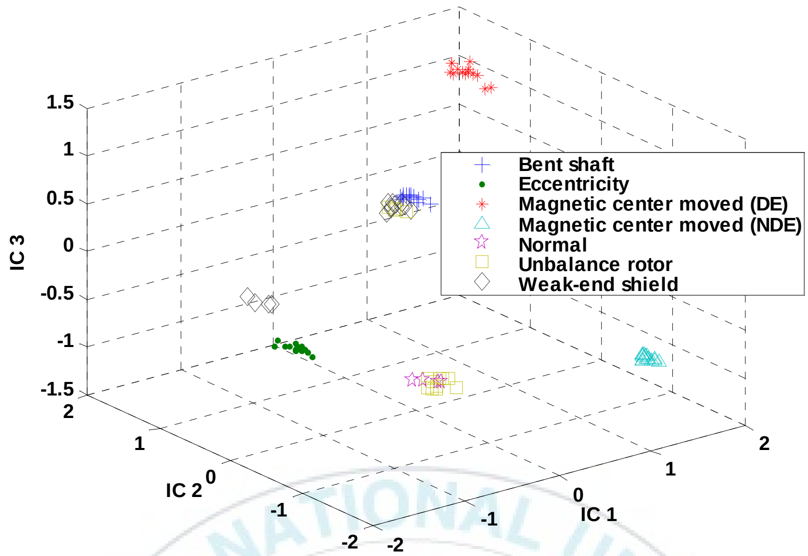
(a) Principle components



(b) Independent components



(c) Kernel principle components



(d) Kernel independent components

Fig. 4.33 The first three principal and independent components.

5.4.4. Training and Classification

The SVM based multi-class classification is applied to perform the classification process using one-against-all methods. To solve the SVM problem, Vapnik [27] describe a method which used the projected conjugate gradient algorithm to solve the SVM-QP problem. In this study, SVM-QP was performed to solve the classification problem of SVM. The parameter *C* (bound of the Lagrange multiplier) and (condition parameter for QP method) were 1 and 10^{-7}, respectively.

Wavelet kernel function using Daubechies series was performed in this study. The parameter δ in wavelet kernel refers to number of vanishing moment and is set 4. In the training process, the data set was also trained using RBF kernel function as comparison. The parameter γ for bandwidth RBF kernel was user defined equal to 0.5.

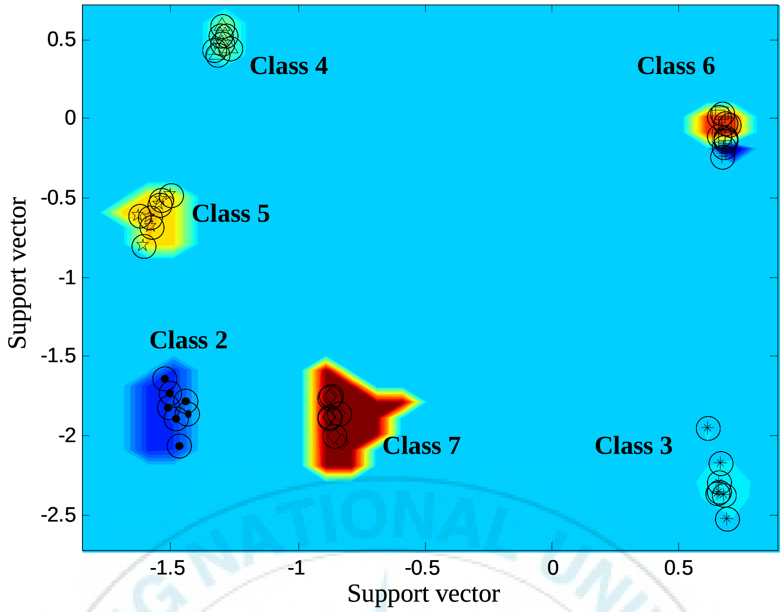
5.4.5. Result and Discussion

The complex separation boundaries are presented in Fig. 4.34 from which the separation of W-SVM can be shown. In these figures, the circle refers to the support vector that states the correct recognition in W-SVM. Each condition of induction motor is well recognized using Daubechies wavelet kernel. In the classification process using W-SVM, each condition of induction motors can be clustered well. The good separation among conditions shows the performance of W-SVM doing recognition of component analysis from vibration signal features.

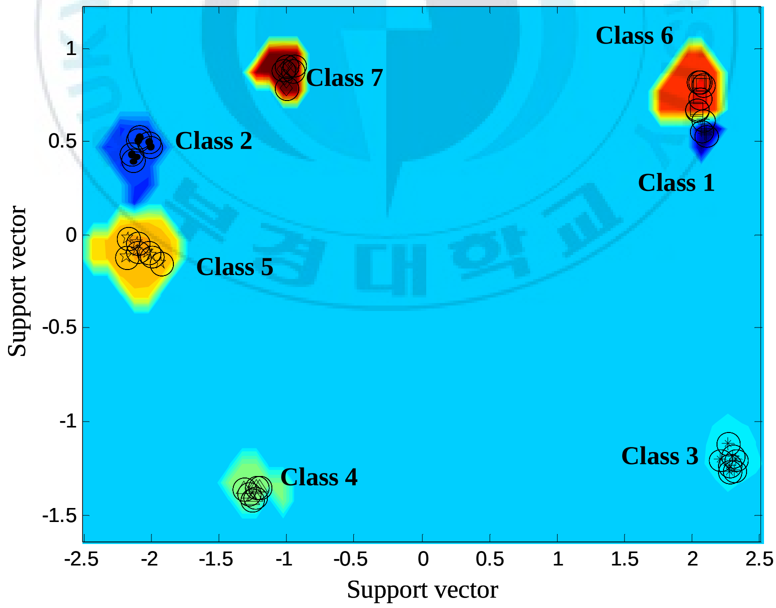
The performance of classification process is summarized in Table 4.17. All data set come from component analysis are accurately classified using Daubechies wavelet kernel and SVM and reached accuracy 100% in training and testing, respectively. SVM using RBF kernel function with kernel width $\gamma = 0.5$ is also performed in classification for comparison with Daubechies wavelet kernel. The results show that the performance of W-SVM is similar to SVM using RBF kernel function, those are 100% in accuracy of training and testing, respectively. In the case of number support vectors, SVM with RBF kernel function needs lower than W-SVM except kernel PCA.

Table 4.17 Results of classification

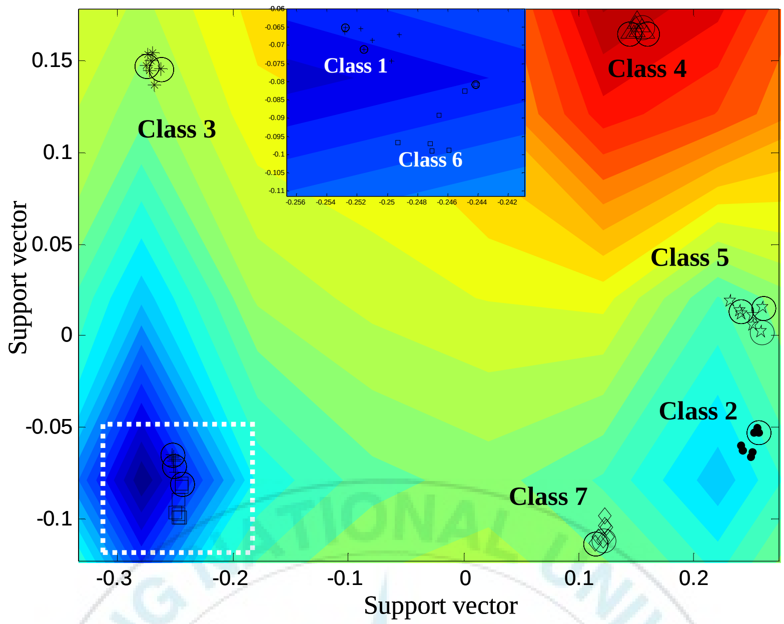
Kernel	Accuracy (Train/Test), %				Number of SVs			
	IC	PC	Kernel IC	Kernel PC	IC	PC	Kernel IC	Kernel PC
Daubechies	100/100	100/100	100/100	100/100	35	39	39	17
RBF-Gaussian (γ= 0.5)	100/100	100/100	100/100	100/100	22	22	25	33



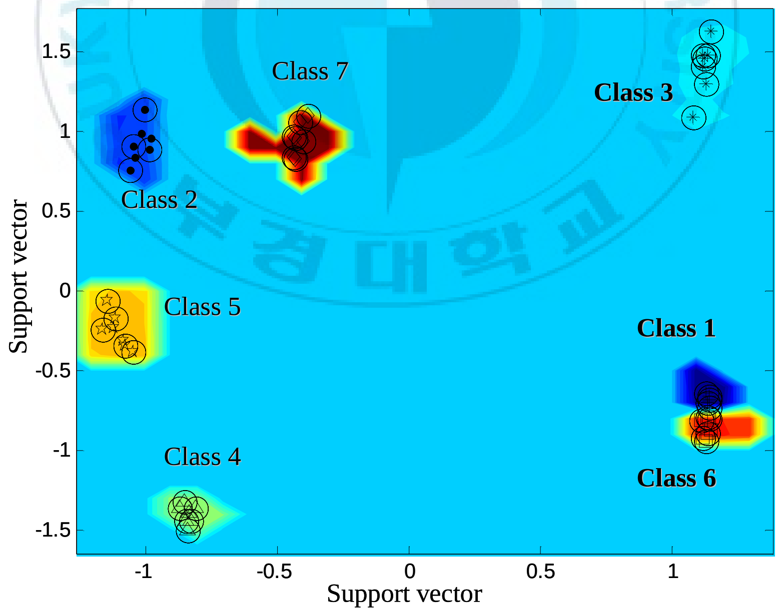
(a) Daubechies kernel with PC data



(b) Daubechies kernel with IC data



(c) Daubechies kernel with kernel PC data



(d) Daubechies kernel with kernel IC data

Fig. 4.34 Separation boundaries of W-SVM.

Based on previous case studies, there are advantages and disadvantages in SVM classification that can be studied as follows [32,33]:

§ Advantages

1. SVM can well learn the system based on training process using small set number of data.
2. SVM has good generalization ability so that it can produce accurate performance in classification when the system is tested.
3. SVM has ability that can be independent of the dimensionality of the feature space.
4. In SVM, embedding structural risk minimization (SRM) principle can minimize the upper bound on the generalization error.

§ Disadvantages

1. Basically, SVM is developed for binary classification. Recently, SVM has been modified to be able to solve multi-class classification problem using special strategy i.e., one-against-one, one-against-all, etc. However, each strategy has merit and demerit that is still open to be investigated for improvement.
2. SVM has problem in using kernel function. There are no exact ways to select a proper kernel function for special case.
3. Proper parameter tuning is still a problem in SVM. Kernel parameters selection is sometimes take much CPU time.

6. Conclusion

The excellent and capability of support vector machine (SVM) in fault diagnosis of induction motor has been explored in this chapter. Four case studies have been presented to validate the proposed method.

In the first case study, we applied the combination of ICA and SVM for intelligent fault diagnosis of induction motor. ICA and PCA were successfully applied for feature extraction process; however, the clustering feature using ICA is better than PCA does. The feature extraction is one important step in fault classification process because it can remove the redundancy and avoid the curse of dimensionality phenomenon. After feature extraction, we performed feature selection process to remove irrelevant and useless feature. The distance evaluation technique was employed due to its simple and reliability. SVM based multi-class classification is applied to do faults classification process. To show the importance of feature extraction and kernel parameters selection, we trained the SVMs onto the data input without and with feature extraction, and then followed by kernel parameters selection. The results show that using ICA feature extraction and combining kernel parameters selection gave the best faults classification. According to this result, the combination of ICA and SVM can serve as a promising alternative for intelligent faults diagnosis in the future.

Second case study discussed the application of nonlinear feature extraction and SVM for faults diagnosis. In this method, we employed and adopted kernel trick for mapping the data features into high dimensional space. Moreover, ICA has formulated in the kernel-inducing feature space and a two-phase kernel algorithm that is kernel PCA plus ICA is developed. Kernel PCA 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 kernel PCA whitened space and determined the mutual components. The effectiveness of nonlinear feature extraction is verified using data feature parameters of induction motor. Feature extraction using linear technique is also introduced to compare with nonlinear one. The result shows that kernel ICA outperforms kernel PCA in clustering based on the investigation of average of Euclidean distance. According to the result, the application of nonlinear feature extraction and SVM can serve as a promising alternative for intelligent faults diagnosis in the future.

In the third and fourth case study, a new method of nonlinear kernel based on wavelet (W−SVM) is introduced. The kernel function transforms the data into higher dimensional space in order to make it possible to perform the separation process. Feature reduction and extraction using component analysis via PCA and KPCA are highlighted. The performance of W−SVM is validated by applying it to faults detection and classification of induction motor based on start-up transient current and vibration signals. The results show that W−SVM is well performed and reached high accuracy in training and testing process based on experimental work. However, a proper preprocessing for the transient current signal is needed to improve emerging the salient differences between conditions in induction motors. Introducing nonlinear kernel using wavelets is believed to improve significantly the SVM research fields.

References

[1] Wildi, T., Electrical Machines, Drives and Power Systems, Prentice Hall Int., 1997.

[2] P.F. Allbrecht, J.C. Appiarius, R.M. McCoy, E.L. Owen, D.K. Sharma, Assessment of the reliability of motors in utility applications, IEEE Transaction on Energy Conversion EC-1 (1) (1986) 39-46.

[3] Benbouzid, M., A review of induction motor signature analysis as a medium for faults detection., IEEE Transactions on Industrial Electronics 4 (5) (2000) 984-993.

[4] Obaid R. et al., Stator current analysis for bearing damage detection in induction motors, in: SDEMPED 2003, Symposium on Diagnostics for Electric Machines, Power Electronics and Drives Atlanta, GA, USA, 2003.

[5] Eschmann P.et al, Ball and Roller Bearings: Their theory, design, and application, London: K. G. Heyden, 1958.

[6] R. Schiltz, Forcing frequency identification of rolling element bearings, Sound Vibration (1990) 16-19.

[7] R. R. Schoen et al., Motor bearing damage detection using stator current monitoring, IEEE Transaction Industrial Application 31 (1995) 1274-1279.

[8] M. A. Cash, T. G. Habetler, Insulation failure prediction in AC machines using line-neutral voltage, IEEE Transaction on Industrial Application 34 (1998) 1234-1239.

[9] M. A. Cash, T. G. Habetler, Insulation failure prediction in stator windings of ASD-driven induction machines using standard deviation of line currents, in Proc. IEEE-IAS Annual Meeting, 1998, pp. 299-303.

[10] J. Penman, H.G. Sedding, B. A. Lloyd, W. T. Fink, Detection and location of interturn short circuits in stator windings of operating motors, IEEE Transaction on Energy Conversion 9 (1994) 652-658.

[11] A. H. Bonnet, G.C. Soukup, Cause analysis of stator and rotor failures in three-phase squirrel-cage induction motors, IEEE Transaction on Industrial Application 28(4) (1992) 921-937.

[12] A, Bellini, F. Filppeti, G. Franceschini, C. Tassoni, Closed loop control impact on the diagnosis of induction motor faults, IEEE Transaction on Industrial Application 36 (2000) 1318-1329,

[13] W. T. Thomson, M. Fenger, Industrial application of current signature analysis to diagnosis faults in 3-phase squirrel cage induction motors, in Proceeding of the IEEE Pulp and Paper Ind. Tech. Conf., 2000, pp. 205-211.

[14] A, Bellini, F. Filppeti, G. Franceschini, C. Tassoni, G. B. Kliman, Quantitave of induction motor broken bars by means of electrical signature analysis, IEEE Transaction on Industrial Application 37 (2001) 1248-1255.

[15] N. M. Elkasabgy, A. R. Eastham, G. E. Dawson, Detection of broken bars

in the cage rotor on induction machine, IEEE Transaction on Industrial Application 28 (1992) 165-171.

[16] S. M. A. Cruz, A. J. M. Cardoso, Further developments on the use of the synchronous reference frame current Park's vector approach, in Proc. IEEE Int. Symp. Diagnostic Elect. Mach., Power Electron. Drives, (2001) 467-472.

[17] D.G. Dorrel, W. T. Thomson, S. Roach, Analysis of airgap flux, current and vibration signals as a function of the combination of static and dynamic air gap eccentricity in 3-pahse induction motors, IEEE Transaction on Industrial Application 33 (1997) 24-34.

[18] S. Nandi, A. Ahmed, H.A. Toliyat, Detection of rotor slot and other eccentricity related harmonics in three-phase induction motor with different rotor cages, IEEE Transaction on Energy Conversion 16 (2001) 253-260.

[19] R. R. Obaid, T. G. Habetler, D. J. Gritter, A simplified technique for detecting mechanical fault using stator current in small induction motor, in Prc. IEEE-IAS Annu. Meeting 1, 2000, pp. 479-483.

[20] B. S. Yang, T. Han, W. W. Hwang, Fault diagnosis of rotating machinery based on multi-class support vector machines, Journal of Mechanical science and Technology 19 (2005) 845-858.

[21] B. S. Yang, T. Han, J. L, An, ART–KOHONEN neural network for faults diagnosis of rotating machinery, Mechanical System and Signal Processing (18) (2004) 645-657.

[22] B.S. Yang, D.S. Lim, J. L. An, Vibration diagnostic system of rotating machinery using artificial neural network and wavelet transform, in Proceeding of 13th International Congress on COMADEM, Houston, 2000, pp. 12-20.

[23] C. W. Hsu, C. C. Chang, C. J. Lin, A practical guide to support vector classification, in Technical Report, Department of Computer Science and

Information Engineering, National Taiwan University, (2005) Available at: http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.

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

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

[26] C. Liu, H. Wechsler, Independent component analysis of Gabor features for face recognition, IEEE Transaction on Neural Networks 14(4) (2003) 919-928.

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

[28] 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).

[29] S. Nandi, H.A. Toliyat, Condition monitoring and fault diagnosis of electrical machines- A review, IEEE–IAS Annual Meeting 1 (1999) 197-204.

[30] M. Haji, H.A. Toliyat, Pattern recognition- A technique for induction machines rotor fault detection, Proceeding of the IEEE–IEMDC, Boston, 2001, pp. 899-904.

[31] Curve Fitting Toolbox for Use with MATLAB. The MathWorks, Inc., Natick, USA, 2002.

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

[33] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20(3) (1995) 273-297.

V. Feasibility of SVM for Machine Prognosis System

1. Introduction

A real machine prognosis system is very important to predict the degradation condition and fault propagation trend in machines before a fault reaches in critical level. It also can produce the early alarm and warning before catastrophic condition occurred. Machine condition prognosis means the use of available (current or previous) observations to predict upcoming states of machine [1]. Compared to fault diagnosis, the papers that concern with prognosis are much fewer. The most widely used of prognosis system is to predict the remaining useful life (RUL) that predicts how much time is left before serious failures occur (one or more faults) based on the current and past conditions of machine. The other method of prognosis is addressed to predict a chance that machines operate without a fault or failure up to some future time until next inspection interval based on the current conditions and past operation profile. This is, actually, more desirable to be a reference for maintenance engineer to determine whether the next inspection interval is proper or not.

Many temporal patterns can be used for machine condition prognosis, such as vibration features and debris properties of lubrication oil. The vibration based monitoring; however, is a well-accepted approach due to the ease of measurement and analysis. Several studies based on vibration techniques have been reported in time-series prediction incorporated with classical approach, such as autoregressive (AR) model, autoregressive moving average (ARMAX) model, bilinear model and multivariate adaptive regression splines. However, the difficulties are found when these models will be applied in predicting of dynamic response of complex system.

Recently, many researchers tend to apply artificial intelligence (AI) techniques due to the ability to be flexible model predictors which can be automatically built by training process without the need for identification of model structures and parameters. The most widely used of AI techniques for forecasting are neural networks (NNs) and fuzzy system. Zhang and Ganesan [2] used self-organizing neural networks for multivariable trending of the fault to estimate the residual life of a bearing system. Wang and Vachstevanos [3] applied dynamic wavelet neural networks to predict the fault propagation and estimate the RUL as the time left before the fault reach a given value. Yam et al. [4] applied a recurrent neural networks for predicting the machine condition trend. Wang et al. [1] compared the result of applying recurrent neural networks and neuro-fuzzy inference system to predict the fault damage propagation trend.

Support vector machine (SVM), introduced originally by Vapnik [5] is one of machine learning methods and AI techniques which has been rapidly developed and applied for classification and regression problem [6,7]. SVM is quite satisfying from a theoretical point of view and can lead to great potential and superior performance in practical applications. This is largely due to the structural risk minimization (SRM) principle in SVM, which has greater generalization ability and is superior to the empirical risk minimization (ERM) principle as adopted in neural networks. Furthermore, SVM is adaptive to complex system and robust in dealing with nonlinear data.

Recently, the application of SVM to time-series prediction, called support vector regression (SVR), has shown many breakthroughs and plausible performance, such as travel-time prediction [8], wind speed prediction [9], electricity load forecasting [10], water lake prediction [11], etc. Since there are much evidence from previous research results of time-varying application with

SVR prediction, it motivates our research in using SVR for machines prognosis system modeling.

In present chapter, SVR is applied to predict time-series of failure trending data of machines. The aims of this study are to investigate the feasibility and to evaluate the performance and reliability of SVR in failure trending data prediction, and also to develop a reliable prognosis system for machines condition prediction.

2. Description of Selected Model

2.1. Support Vector Regression (SVR)

Recall the linear equation of SVM expressed in Eq. (3.31), it can be express in the form

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b \tag{5.1}$$

where $\langle \, , \rangle$ denotes the dot product in R^n.

Flatness in the case of Eq. (5.1) means the one seeks small **w**. One way to ensure this is to minimize the Euclidean norm, i.e. $\|\mathbf{w}\|^2$. Formally, the problem of Eq. (5.1) can be written as convex optimization problem by requiring

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\|\mathbf{w}\|^2 \\ &\text{subject to} \quad \begin{cases} \mathbf{y}_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b \le \varepsilon \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b - \mathbf{y}_i \le \varepsilon \end{cases} \end{aligned} \tag{5.2}$$

The tacit assumption in Eq. (5.2) is that such a function $f(\mathbf{x})$ actually exists that approximately all pairs (x_i,y_i) with ε precision, or in other words, that the convex optimization is feasible. However, this may not be the case, or we also may want to allow some errors. Analogously to the soft margin in Vapnik [12], one can introduce slack variables ξ_i, ξ_i^* to cope with otherwise infeasible constraints to optimization Eq. (5.2). Hence, we present the formulation stated in [5].

$$\text{minimize} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)$$

$$\text{subject to} \quad \begin{cases} \mathbf{y}_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b \le \varepsilon + \xi_i \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b - \mathbf{y}_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{5.3}$$

The constant $C > 0$ determines the trade off between the flatness of $f(x)$ and the amount up to which deviations larger than ε tolerated. The formulation above corresponds to dealing with a so called ε-insensitive loss function $|\xi|_\varepsilon$ described by

$$|\xi|_\varepsilon = \begin{cases} 0 & \text{if } |\xi| \le \varepsilon \\ |\xi| - \varepsilon & \text{otherwise} \end{cases} \tag{5.4}$$

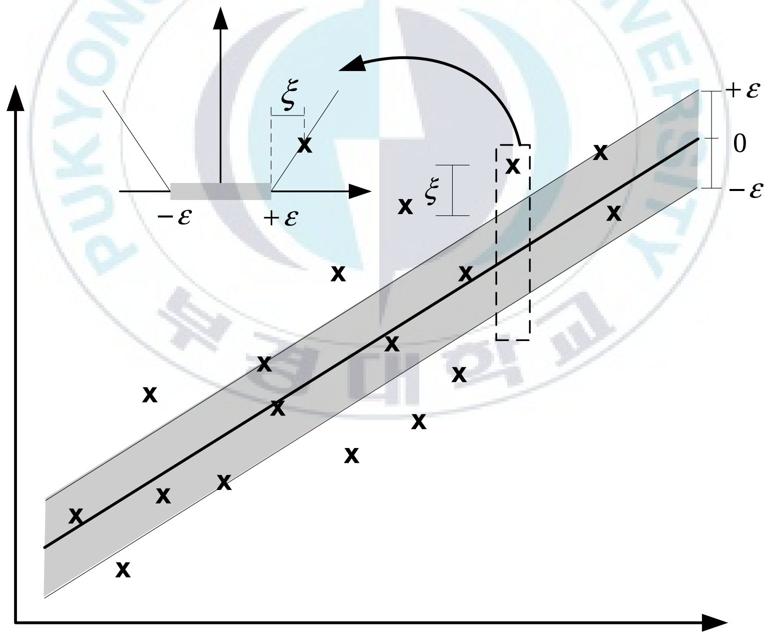


Fig. 5.1 The soft margin loss setting for linear SVR.

Fig. 5.1 depicts situation graphically. Only the points outside the shaded region contribute to the cost insofar, as the deviation are penalized in a linear fashion. It turns out that the optimization problem in Eq. (5.3) can be solved more easily in its dual formulation. The dual function provides the key for extending support vector machine to nonlinear functions.

The calculation can be simplified by converting into the equivalent *Lagrangian* dual problem, which will be

$$
\begin{aligned}
L(\mathbf{w},b,\alpha) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}(\xi_i+\xi_i^*) - \sum_{i=1}^{l}\alpha_i(\varepsilon+\xi_i-\mathbf{y}_i+\langle\mathbf{w},\mathbf{x}_i\rangle+b) \\
-\sum_{i=1}^{l}\alpha_i^*(\varepsilon+\xi_i^*-\mathbf{y}_i+\langle\mathbf{w},\mathbf{x}_i\rangle-b) - \sum_{i=1}^{l}(\eta_i\xi_i+\eta_i^*\xi_i^*)
\end{aligned}
\tag{5.5}
$$

Then, the task is minimizing Eq. (5.3) with respect to primal variables (**w**, b, ξ_i, ξ_i^*) have to famish for optitimality.

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^{l}(\alpha_i^*-\alpha_i)\mathbf{x}_i = 0 \tag{5.6}$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{l}(\alpha_i^*-\alpha_i) = 0 \tag{5.7}$$

$$\frac{\partial L}{\partial \xi_i^{(*)}} = C - \alpha_i^{(*)} - \eta_i^{(*)} = 0 \tag{5.8}$$

Substituting Eqs. (5.6), (5.7), (5.8) into Eq. (5.5) yields the dual optimization problem.

$$
\begin{aligned}
&\text{minimize} \quad \left\{\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)\langle\mathbf{x}_i,\mathbf{x}_j\rangle - \varepsilon\sum_{i=1}^{l}(\alpha_i+\alpha_i^*) + \sum_{i=1}^{l}\mathbf{y}_i(\alpha_i-\alpha_i^*)\right. \\
&\text{subject to} \quad \begin{cases}\sum_{i=1}^{l}(\alpha_i-\alpha_i^*) = 0 \\ \alpha_i,\alpha_i^* \in [0,C]\end{cases}
\end{aligned}
\tag{5.9}
$$

In deriving Eq. (5.9), the dual variables η_i, η_i^* have eliminated through condition Eq. (5.8), as the variables did not appear in the dual objectives function anymore but only were present in the dual feasibility conditions. Eq. (5.6) can be rewritten as follows

$$\mathbf{w} = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*)\mathbf{x}_i \tag{5.10}$$

And therefore Eq. (5.1) can be expressed as

$$f(\mathbf{x}) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*)\langle \mathbf{x}_i, \mathbf{x} \rangle + b \tag{5.11}$$

This is so called support vector expansion, i.e. can be completely described as a linear combination of the training patterns $\mathbf{x}_i$. The Lagrange multipliers α_i and α_i^* represent solutions to the above quadratic problem, which act as forces pushing predictions toward target value $\mathbf{y}_i$. Only the nonzero values of the Lagrange multipliers in Eq. (5.9) are useful in forecasting the regression line.

In Eq. (5.11), the dot product of $\langle \mathbf{x}_i, \mathbf{x} \rangle$ can be replaced with function $K(\mathbf{x}_i,\mathbf{x})$ known as the kernel function. Kernel functions enable the dot product to be performed in high-dimensional feature space using low dimensional space data input without knowing the transformation. All kernel function must satisfy Mercer's condition [13] that corresponds to the inner product of some feature space. The RBF is commonly used as kernel for regression

$$K(\mathbf{x}_i, \mathbf{x}) = \exp\{-\gamma |\mathbf{x}_i - \mathbf{x}|^2\} \tag{5.12}$$

For the variable b, it can be computed by applying the Karush-Kuhn-Tucker (KKT) condition that, in this case, imply that the product of the Lagrange multipliers and constrains has to equal to 0

$$\alpha_i(\varepsilon + \xi_i - \mathbf{y}_i + \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0 \tag{5.13}$$

$$\alpha_i^*(\varepsilon + \xi_i^* - \mathbf{y}_i + \langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 0 \tag{5.14}$$

and

$$(C - \alpha_i)\xi_i = 0 \tag{5.15}$$

$$(C - \alpha_i^*)\xi_i^* = 0 \tag{5.16}$$

Since α_i, $\alpha_i^*= 0$, and $\xi_i^* = 0$ for $\alpha_i^* \in (0,C)$ b can be computed as

$$b = \mathbf{y}_i - (\mathbf{w}_i, \mathbf{x}_i) - \varepsilon \qquad \text{for} \quad \alpha_i \in (0,C) \tag{5.17}$$

$$b = \mathbf{y}_i - (\mathbf{w}_i, \mathbf{x}_i) + \varepsilon \qquad \text{for} \quad \alpha_i^* \in (0,C) \tag{5.18}$$

2.2. Prediction Method

Let $\{x(t)\}$, $t = 1, \ldots, T$, be a timer series that was generated by dynamical system. For convenience, consider $x(t)$ to be scalar, but note that the treatment of multi-scalar time series is straightforward. By assuming that $\{x(t)\}$ is a projection of a dynamics operating in a high-dimensional space. If the dynamics is deterministic, the prediction of time series can be performing by reconstructing the state space. The way for reconstruction was introduced by Packard et al. [14] and mathematically analyzed by Takens [15]. A state vector is defined as

$$\mathbf{x}_t = (x(t), x(t-\tau), \ldots, x(t-(d-1)\tau)) \tag{5.19}$$

with time-delay τ and embedding dimension d. If the dynamics runs on an attractor of dimension D a necessary condition for determining $\mathbf{x}_t$ is

$$d \geq D \tag{5.20}$$

If the embedding dimension is big enough, such that $\mathbf{x}_t$ unambiguously describes the state of the system at time t then there exists an equation for points on the attractor, which is of the form

$$\mathbf{x}(t+p) = f^*(\mathbf{x}_t) \tag{5.21}$$

In this equation, f^* is a function that allows to predict future values of the time series $\{x(t)\}$ given past values, with p being the prediction horizon.

Regression technique can therefore be used to estimate the prediction function on the basis of time-delay coordinates according to Eq. (5.19).

3. Methodology

In present study, SVR is applied to forecast the degradation condition trends in rotating machinery. Usually, when a fault induces in rotating machine, at same time the degradation condition will be occurred. The degradation condition of machine can be indicated by the increasing of vibration level in associated machine elements. Vibration-based machine fault prognosis is to use available vibration symptom to predict upcoming states of the fault propagation and degradation condition trend by monitoring one or more parameters. In this section, the methodology of machine fault prognosis is described in Fig. 5.2 as follow

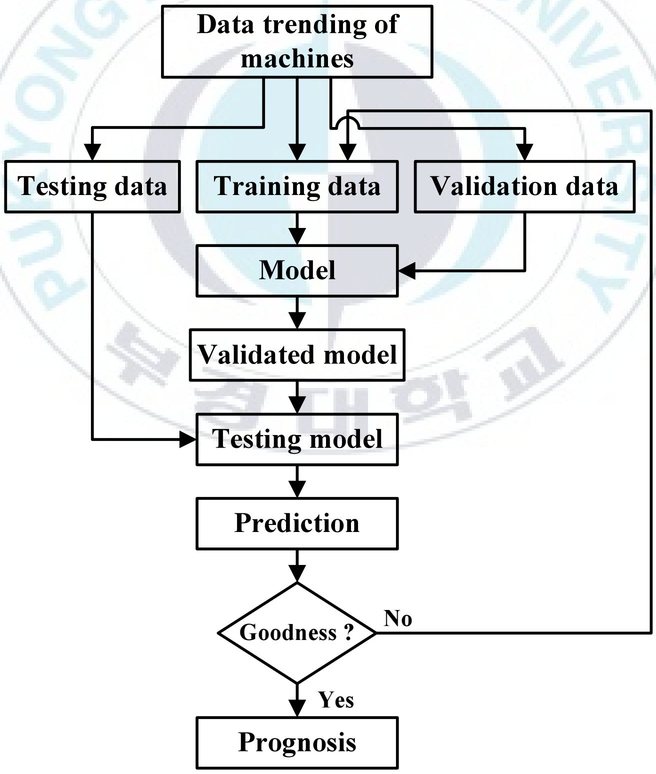


Fig. 5.2 Flowchart of prognosis system using SVR.

1. *Data acquisition*: the data to be used is data trending of machine based on vibration signal which contains data histories of machine until faults occurred.
2. *Data sectioning*: the trending data is divided into three parts: training data, validation data and testing data. The training and validation data are used to build the model for machine fault prognosis system, while the testing data is used to test the validated model. After model validation, the tested model will be obtained.
3. *Prediction*: the tested model is used to predict the future data that is never used for training and validation. The goodness of prediction result is measured by performance measures e.g., root-means square error (*RMSE*) and correlation coefficient *R*.
4. *Prognosis system*: it is obtained if the prediction is successful and passed the user defined criterion of performance measures.

4. Data Benchmarking

In this section, the prediction performance of SVM predictors is evaluated using two typical data sets: a sunspot activity record and a Mackey-Glass equation data series. These are benchmark data set in time-series prediction research due to their specific natures such as nonlinear, non-Gaussian, and non-stationary for the former, and chaotic, non-periodic, and non-convergence for the latter.

The verification performance statistic, such as the root-mean square error (*RMSE*) and correlation coefficient (*R*) are used to examine the system. *RMSE* provides a general illustration of the overall accuracy of the prediction s they show the global goodness of fit, given as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \tag{5.22}$$

where N represents the total number of data points in the test set; y represents the observed value and $\hat{y}$ represents the predicted value. The correlation statistic coefficient (R) measure the linear correlation between the actual and predicted value, it can be calculated as

$$R_{y,\hat{y}} = \frac{Cov(y,\hat{y})}{\sigma_y \sigma_{\hat{y}}} \tag{5.23}$$

where R is correlation coefficient and Cov (y, $\hat{y}$) is covariance between observed and predicted values, which can be calculated as follows

$$Cov(y,\hat{y}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}) \tag{5.24}$$

where $\bar{y}$ is the mean of the observed value and $\bar{\hat{y}}$ is the mean of predicted value. The standard deviation of the observed and predicted values, σ_y and $\sigma_{\hat{y}}$, respectively, can be calculated as

$$\sigma_y = \left(\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2\right)^{1/2} \tag{5.25}$$

$$\sigma_{\hat{y}} = \left(\frac{1}{N-1}\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2\right)^{1/2} \tag{5.26}$$

4.1. Sunspot Data

The sunspot data set can be considered as a nonlinear and non-stationary data. It has served as a benchmark and been well studied in previous literature [14,15]. The available data set used here contains the sunspot activity record for the period from years 1700 to 2005. This data, displayed in Fig.5.3, can be downloaded from Online Sunspot Data Archive, SIDC, RWC Belgium World Data Center (http://sidc.oma.be/index.php3).

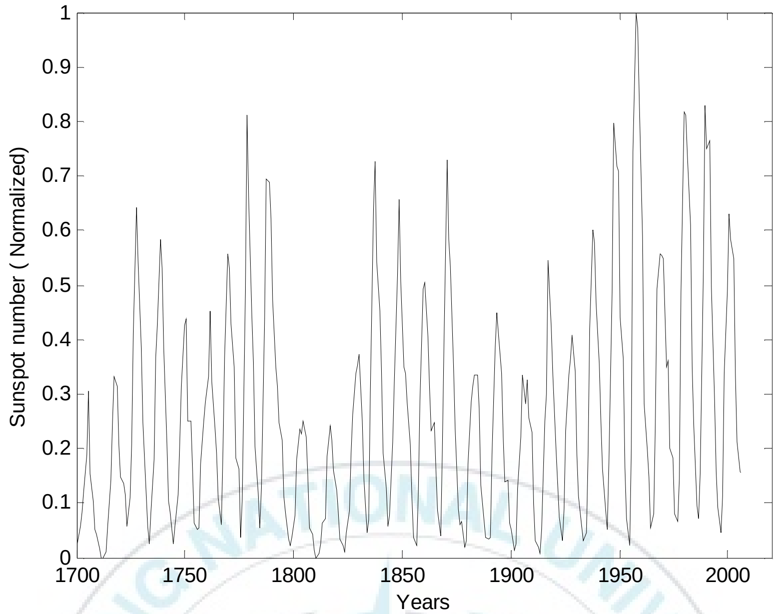


Fig. 5.3 Sunspot activity record from years of 1700 to 2005.

The first 249 sample (year 1700-1948) are used to train and to validate the system, while the remaining data pairs are used to test the identified models. For this training and validation, 5-fold cross-validation is performed to select the proper parameters of RBF kernel function in SVM, those are kernel width (γ) and regularization parameter (C). In the selection process, the parameters which give minimum cross-validation error are selected and used for time series prediction using SVM. Cross-validation process gives proper kernel parameters $\gamma = 2$ and $C = 1$, and ε-insensitive loss function is user defined equal to 0.001. The performance of validation process is presented in Fig. 5.4. Cross-validation selects randomly the points from data set for training and validating the system. After training completed, the support vectors, weights and bias are resulted and used to validate the system. Validation process can indicate the quality and performance of the system being established. Fig. 5.4 shows that the system has good performance according to the performance measure using *RMSE* and correlation coefficient (*R*). *RMSE* and *R* reached 0.0188 and 0.98, respectively, are satisfied. It means the process of learning and validating of SVM to establish the prediction system is successful and the model is generated.

Then, the testing process should be performed to test the model using independent data set that is never used in training and validating process. The support vectors, weights and bias which already saved are employed to test the performance the model.

Fig. 5.5 demonstrate the prediction result of testing data using support vector regression (SVR), examining this graph using *RMSE* and correlation coefficient *R*, the SVR provides a reasonably well prediction performance.

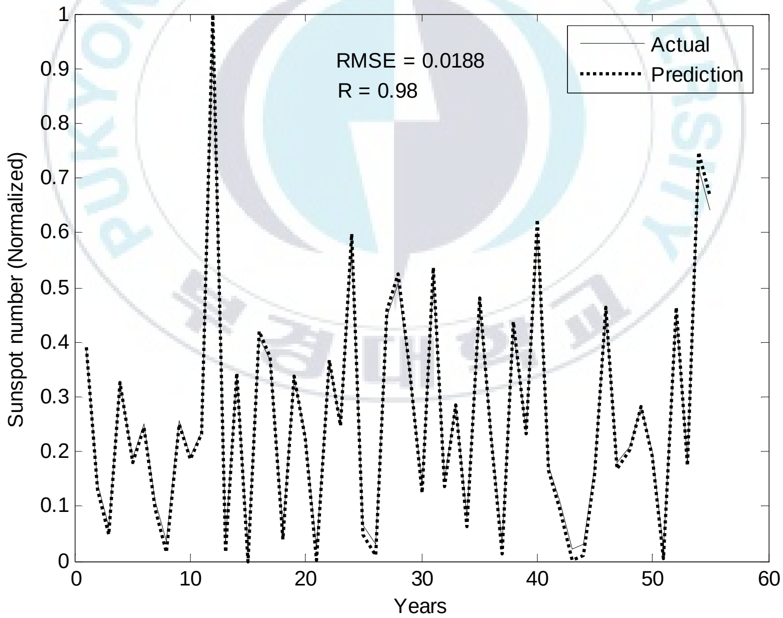


Fig. 5.4 Model validation.

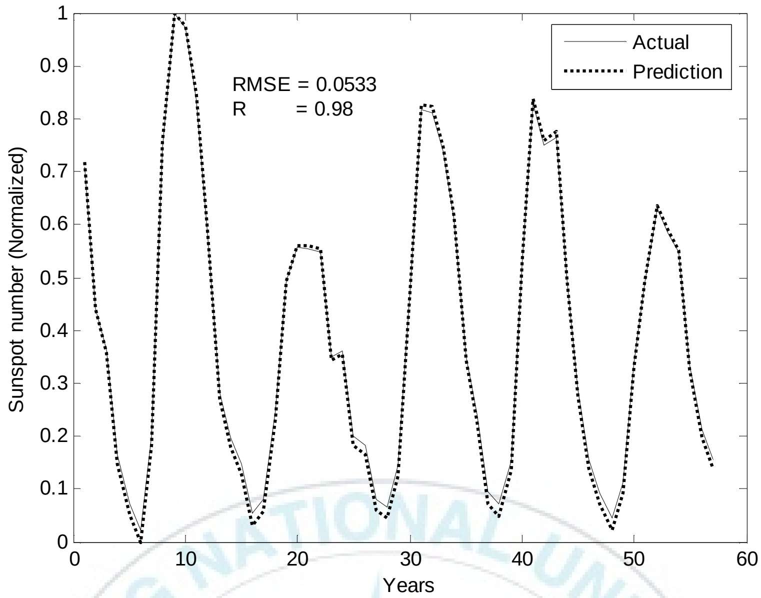


Fig. 5.5 Prediction of sunspot data using SVR.

4.2. Mackey Glass Data

Mackey-Glass (MG) differential delay equation [16] was first proposed for modeling white blood cell production in human bodies, which defined as

$$\frac{dx(t)}{dt} = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t) \tag{5.27}$$

This time series is chaotic and so there is no clearly defined period. The series will not converge and diverge and the trajectory is highly sensitive to initial conditions. This is a frequently used of benchmark problem in the neural network and fuzzy modeling research communities. The initial condition used is $x(0) = 1.2$, $\tau = 17$, and $x(t) = 0$ for $t < 0$, 1,201 data are selected then normalized and plotted in Fig. 5.6.

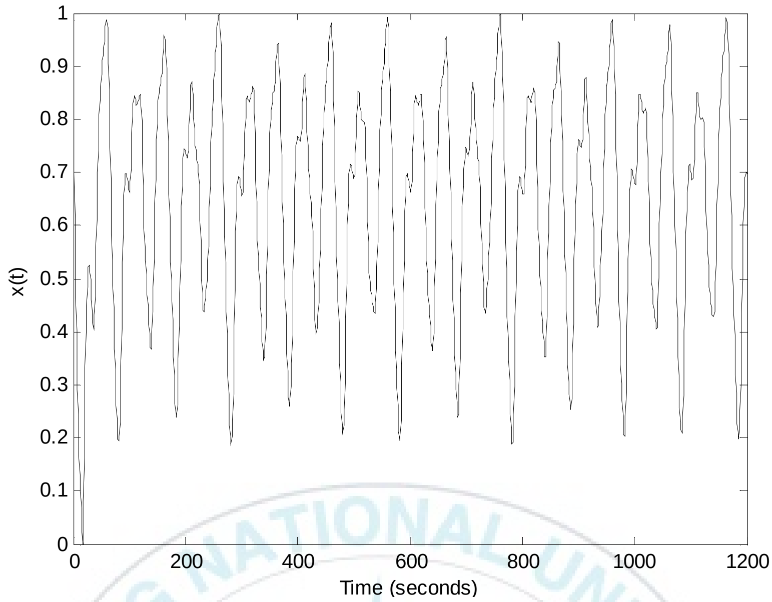


Fig. 5.6 Mackey-Glass differential equation.

The first 800 samples are used for training and validation the system, whereas the remaining samples for testing the model. RBF kernel function is selected when SVM is performed in the training and validation process. 5-fold cross-validation is also employed to select optimal kernel parameters for RBF kernel function and resulted $\gamma = 2$ and $C = 1$. Using ε-insensitive loss function in SVR $\varepsilon = 0.001$, and employing the same way and method such as previous benchmarking, the validating and prediction of Mackey-Glass data is presented in Figs. 5.7 and 5.8. From the comparison between the actual and predicted one, it can be seen that a properly trained SVR can capture the system dynamic behavior accurately and quickly. The performance of system is shown in testing process based on the untrained data (401 samples) of future time values. The training data used is about 66% of total data sets and they can train the system well.

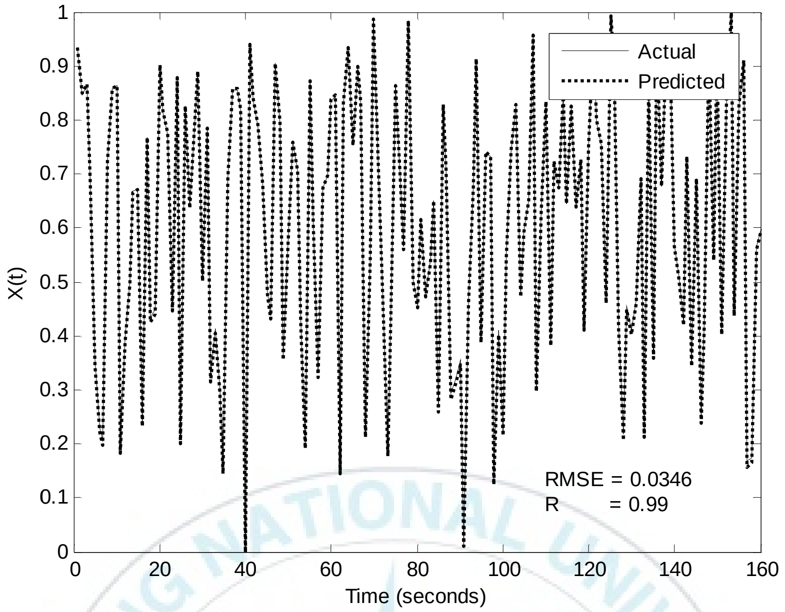


Fig. 5.7 Validation model of system using Mackey-Glass data.

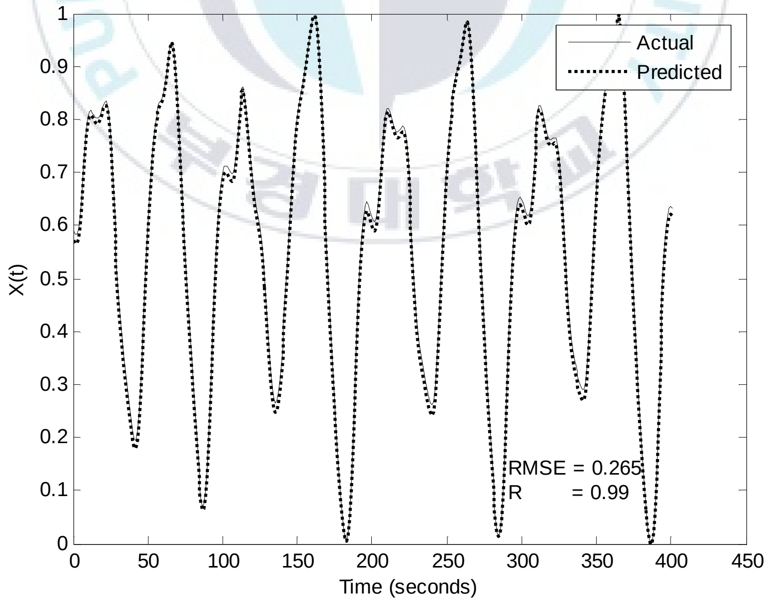


Fig. 5.8 Prediction of Mackey-Glass data using SVM.

According to the results of benchmarking data, SVM based on regression has potential to perform time-series prediction. Even though SVR system has shown great potential in nonlinear and stochastic time-series predictions, there is no known application in real-time machine health condition prognosis. From this reason, machine fault prognosis system can be established based on the excellence performance of regression using SVR.

5. Experiment

The proposed method is validated by applying in real system to predict the trending data of a low methane compressor (Fig. 5.9). This compressor is driven by a motor 440 kW, 6600 volt, 2 poles with operating speed 3565 rpm. The related information of system is summarized in Table 5.1.

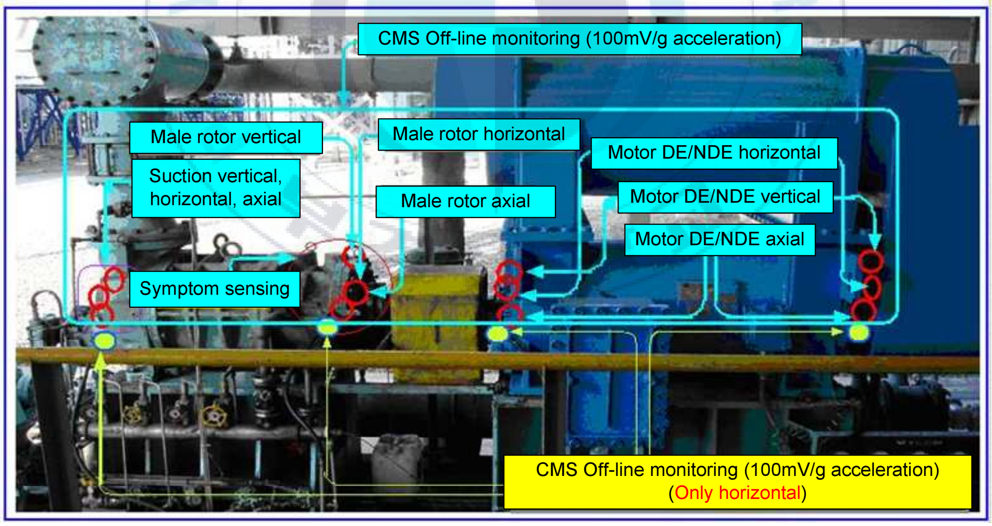


Fig. 5.9 Low methane compressor: wet screw type.

The system consists of two types of condition monitoring those are off-line and on-line system. In off-line system, several vibration sensors are installed in some locations of motor and compressor such as drive-end motor, non drive-end motor, male rotor compressor and suction part of compressor. Each location consists of three directions of measurement: axial, vertical and horizontal. The circle in Fig. 5.9 shows the male of rotor compressor that are symptoms sensing location in this system.

On-line monitoring system consists of acceleration sensor in only horizontal direction of four locations: drive-end motor, non drive-end motor, male rotor compressor and suction part of compressor.

Table 5.1 Description of system

Motor		Compressor	
Voltage	6600 V	Type	320 LUD-MB, Wet Screw (Unload System)
Power Pole	440 kW 2 Pole	Lobe	Male Rotor (4) Female Rotor (6)
Bearing Rpm	NDE/#6216, DE/#6216 3565 rpm	Bearing	Thrust Brg : 7321 BDB Radial Brg : Sleeve

The data used in this experiment are trending data of peak acceleration, envelope acceleration. Trending data were recorded from August 2005 to November 2005 which consists of 400 points. This data contains information of machine history (vibration amplitude) with respect to time sequence which can be regarded as time-series. The proposed method is addressed to predict future condition of vibration amplitude based on the previous state. SVM predictor will learn the characteristic of previous state and save it as weights, bias and support vectors to perform prediction.

6. Result and Discussion

Fig. 5.10 shows the trending data of peak acceleration of low methane compressor. This data consists of 400 points measurement that represents the machine conditions. At the beginning, condition of machine is normal as shown in the figure that the peak acceleration is almost constant until point 300. Over point 300, amplitude of machine drastically increased that means the condition of machine is changed and degradation condition is occurred. Moreover, it indicates that some faults are occurred in the machine that changes the amplitude significantly.

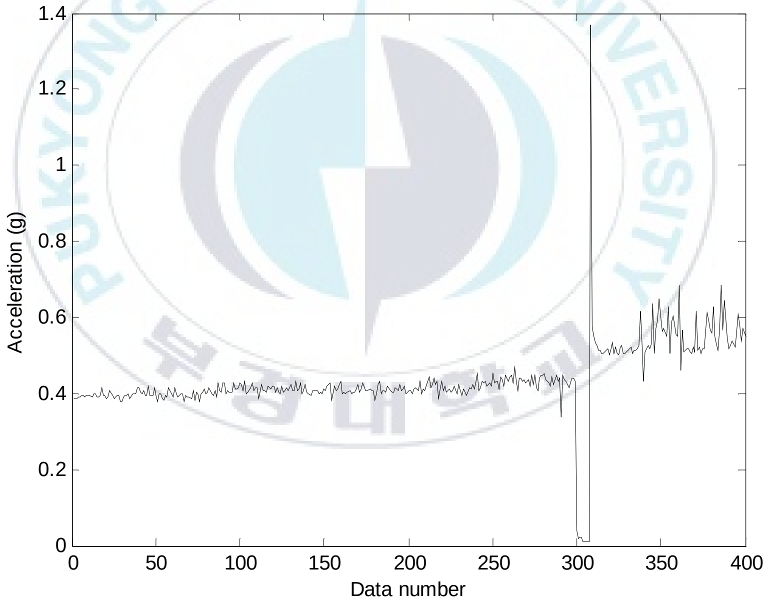


Fig. 5.10 Peak acceleration of low methane compressor.

The proposed method is aimed to predict the future state of machine based on previous conditions. Data from normal state are used to train the proposed system for building the model, and then model will be employed to forecast the future condition of machine. The first 300 data are used for training the system and validating the model, while the remains are used for testing the performance of system.

Training process is performed using 5-fold cross-validation to select the kernel parameters of RBF kernel function. Cross-validation process gives proper kernel parameters $\gamma = 0.25$ and $C = 1$, and ε-insensitive loss function is defined equal to 0.001. The result of model validation is presented in Fig. 5.11 that gives *RMSE* and *R* are 0.035 and 0.70, respectively. The validated model cannot catch the minimum amplitude due to poor of training. However, the error presented by *RMSE* reaches 0.035 is acceptable to be a model although the correlation is small (0.7) because the minimum of amplitude cannot be caught by the model.

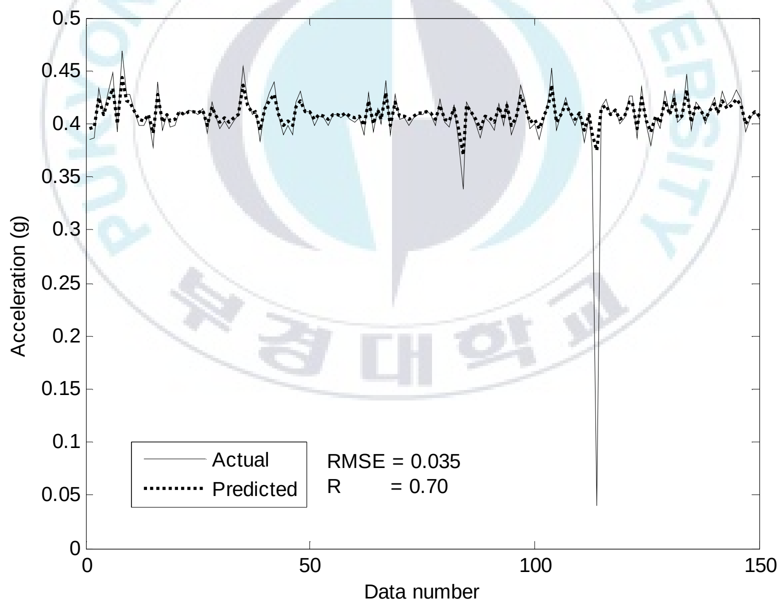


Fig. 5.11 Model validation using peak acceleration data.

Fig. 5.12 depicts the performance of testing using future independent data (100 data points) that is never used in training process. The result seems over prediction that cannot approach actual trending data of peak acceleration. *RMSE* reaches 4.67 is relatively big enough so it may not be a good prediction model. Even though the correlation presented by *R* is 0.7 shows the poor correlation between the predicted value and the actual one, however, the trending of predicted value is relatively similar to the actual data.

The reason why this model has poor performance is the training data do not contain extreme (or relatively close to extreme) value of amplitude. As intelligent system, if the system is experienced or ever taught by relatively close to the extreme value so it might be able to catch the actual values. The other reason is the trending data of peak acceleration is drastically changed when it represents the degradation condition of machine. So the model suffers difficulty to catch the actual values.

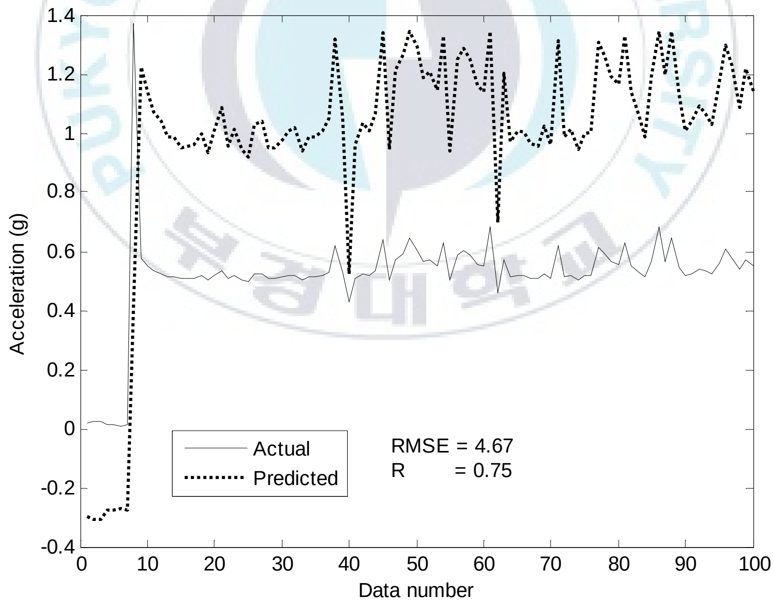


Fig. 5.12 Prediction of peak acceleration data.

Fig. 5.13 demonstrates the trending data of envelope acceleration of low methane compressor. The proposed method is addressed to predict the future state condition of machine based on learning from previous condition. First 300 data are used to train the system for building and validating the model. The remains of 100 data are targeted as actual value that will be predicted by model. The model should predict the maximum value of amplitude that represents the machine degradation or fault occurrence.

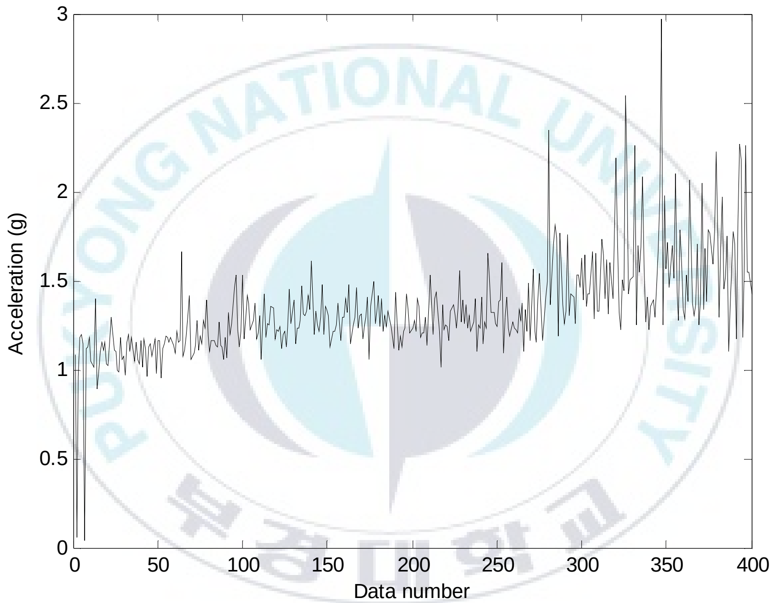


Fig. 5.13 Data trending of envelope acceleration.

SVM is trained by training data using 5-fold cross validation for RBF kernel parameters selection. Cross-validation process gives proper kernel parameters $\gamma = 4.5$ and $C = 1$, and ε-insensitive loss function is defined equal to 0.001. The result of model validation is presented in Fig. 5.14 that gives *RMSE* and *R* are 0.075 and 0.98, respectively. The validated model can catch very well the dynamic system represented by training data. Therefore, the model is acceptable and can be considered to be a model as a predictor for system forecasting.

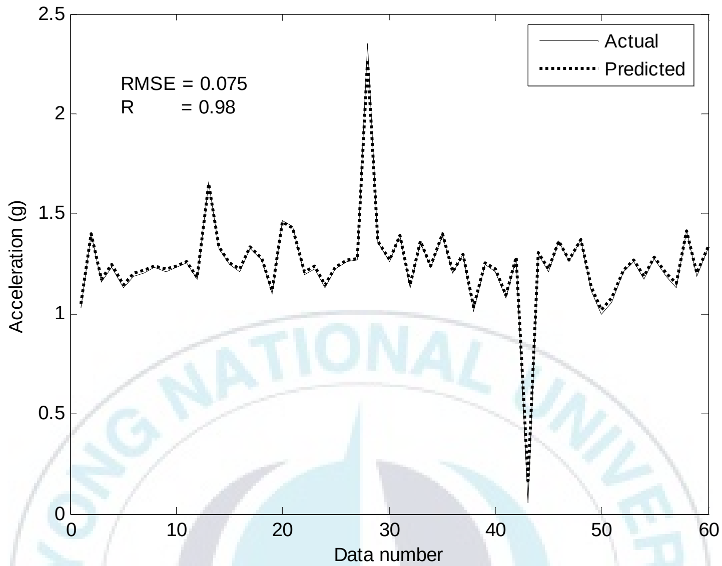


Fig. 5.14 Model validation using envelope acceleration data.

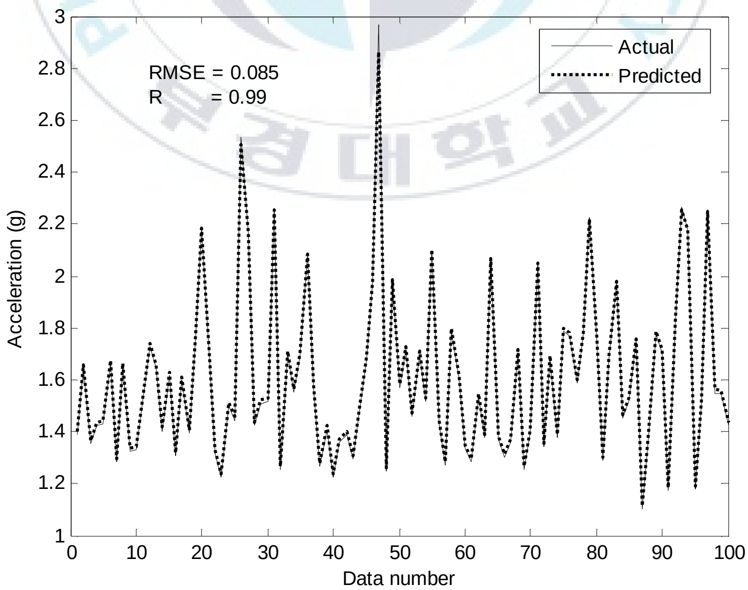


Fig. 5.15 Prediction of envelope acceleration data.

The performance of prediction is depicted in Fig. 5.15 that shows the acceptability of the model. *RMSE* reaches 0.085 is relatively small that means the values of predicted data and actual data are very close. Also, the correlation measure *R* is high, 0.99 which represents the predicted values and the actual one are high-correlated.

In this case, the training process is well performed due to good quality of training data that are close-related among others. It means there are no extreme differences (drastically change) between amplitudes of envelope acceleration. So the prediction using SVM model can perform well.

As general, for SVM regression, there are similar advantages and disadvantages as mentioned in classification task. Moreover, it can be added as follows [11]:

§ Advantages

1. In regression training of SVM, it consists of solving a uniquely solvable quadratic optimization problem, which is much more attractive because it is guaranteed to find a global minimum of the error surface.
2. The use of dual setting in the constrained optimization avoids having to define and compute the parameters of the optimal hyperplane in a data space of possibly high dimensionality.
3. In SVM, the complexity of the machine learning is handled by the support algorithm itself.
4. The computation can be performed efficiently without a large CPU time requirement.

§ Disadvantages

1. Problem of kernel function selection.
2. Problem of proper kernel parameters.

7. Conclusion

Prognosis of machine condition is very important to provide an accurate alarm before fault reaches critical levels so as to prevent machinery performance degradation, malfunction or catastrophic failure. In this chapter, the feasibility of support vector machine (SVM) for prognosis system has been studied. The model predictor is built based on the ability of SVM for regression technique.

Problem benchmarking has been performed using Sunspot data and Mackey-Glass data that are frequently used for benchmarking in machine learning area. These data contains chaotic and complex dynamic behavior, so it is very interesting to apply these data for performance evaluation of the proposed system.

The proposed method is validated by applying it to predict the future state condition of a low methane compressor based on given previous state data. Two cases have been studied using peak acceleration and envelope acceleration. The results show that the proposed method has potential to be a prediction tool for prognosis system based on time-series prediction.

References

[1] W.Q. Wang, M.F. Golnaraghi, F. Ismail, Prognosis of machine health condition using neuro-fuzzy system, Mechanical System and Signal Processing 18 (2004) 813-831.

[2] S. Zhang, R. Ganesan, Multivariable trend analysis using neural networks for intelligent diagnostics of rotating machinery, Transactions of the ASME Journal of Engineering for Gas Turbines and Power 119 (1997) 378–384.

[3] P. Wang, G. Vachtsevanos, Fault prognostics using dynamic wavelet neural networks, AI EDAM-Artificial Intelligence for Engineering Design Analysis and Manufacturing 15 (2001) 349–365.

[4] R.C.M. Yam, P.W. Tse, L. Li, P. Tu, Intelligent predictive decision support system for condition-based maintenance, International Journal of Advanced Manufacturing Technology 17 (2001) 383–391.

[5] V.N. Vapnik, The Nature of Statistical Learning Theory, New York: Springer; 1995.

[6] S.R. Gun, Support vector machine for classification and regression, Tech. rep., University of Southampton, U.K., May 1998.

[7] K.R. Muller, A.J. Smola, B. Scholkopf, J. Kohlmorgen, V. Vapnik, Predicting time series with support vector machine, in Proc. Int. Conf. Artificial Neural Network (ICANN'97), LNCS 1327, 1997, pp. 999-1004.

[8] C. H. Wu, J.M. Ho, D. T. Lee, Travel-time prediction with support vector regression, IEEE Transction on Intelligent Transportation System 5(4) (2004) 276-281.

[9] M.A. Mohandes, T.O. Halawani, S. Rehman, A.A. Hussain, Support vector for wind speed prediction, Renewable Energy 29 (2004) 939-947.

[10] X. Li, C. Sun, D. Gong, Application of support vector machine and similar day method for load foecasting, LNCS 3611 (2005) 602-609.

[11] M. Sajjad Khan, P. Coulibaly, Appliction of support vector machine in lake water level prediction, Journal of Hydrology Engineering 11(3) (2006) 199-205.

[12] C. Cortes, V. Vapnik, Support vector network, M. Leaning 20 (1995) 273-297.

[13] N. Cristianini N.J. Shawe-Taylor, An Introduction to Support Vector Machines, Cambridge: Cambridge University Press; 2000.

[14] A.S. Weigend, B.A, Huberman, D.E. Rumelhart, Predicting the future: a connectionist approach, Int. J. Neural System 1 (1990) 193-209.

[15] M.B. Cotrell, Y. Girard, M. Mangeas, C. Muller, Neural modeling for time series: a statistical stepwise method for weight elimination, IEEE

Transaction on Neural Network 6(6) (1995) 1355-1364.
[16] M. Mackey, L. Glass, Oscilation and chaos in physiological control systems, Science 197 (1977) 423-436.

VI. Conclusions and Future Work

1. Conclusions

In this dissertation, a complete study of fault diagnosis and prognosis by means of support vector machine (SVM) has been deeply studied. Some basic theories including signal analysis in time and frequency domains, which are used for feature representation are reviewed to give preliminary understanding in fault diagnosis procedure. In other word, we used a technique so-called feature-based technique to represent machine conditions. The advantage of this technique is to solve data transfer and data storage problem. Data represented as feature provide better solutions that greatly reduce the requirements of transfer number and storage space, while the information is kept as high as possible.

Feature-based technique involves relative techniques such as feature representation, feature extraction and feature selection. In feature representation, features are calculated from time domain, frequency domain and auto-regression estimation to keep the information at highest level. Large scale features are usually obtained due to multi-sensors used and multi-position of measurement on critical elements that requires much calculation time and degrades accuracy of system. Therefore, feature extractions using linear and nonlinear technique via component analysis are proposed to obtain optimal feature for good fault classification. The linear techniques are principal component analysis (PCA) and independent component analysis (ICA), while nonlinear techniques are employed by introducing kernel function into linear technique.

Support vector machine which is known as new technique in machine learning is highlighted to understand the classification procedure for fault diagnosis system.

In this study, SVM is adopted, redeveloped and combined with feature-based technique to obtain a novel fault diagnosis tool. Moreover, wavelet support vector machine (W-SVM) is introduced to contribute a relatively new technique in classification method used to fault diagnosis routine. Finally, the proposed method is validated using induction motor data to perform fault diagnosis by means of classification strategy in SVM. Several case studies have been done to diagnose fault occurrence in induction motor such as bent rotor, broken rotor bars, bearing fault, mass unbalance, phase unbalance and eccentricity fault. The data used in the experiments are vibration and current data. The results show that the proposed method can perform fault diagnosis well and it can be concluded that the proposed method may serve the fault diagnosis technique in the future.

Prognosis can be defined as the ability to predict accurately and precisely the remaining useful lifetime of a failing machine component or subsystem. In this dissertation, the feasibility of support vector machine (SVM) for prognosis system has been studied. The model predictor is built based on the ability of SVM for regression technique. Problem benchmarking has been performed using Sunspot data and Mackey-Glass data that are frequently used of benchmarking in machine learning area. These data contains chaotic and complex dynamic behavior, so it is very interesting to apply these data for performance evaluation of the proposed system.

The proposed method is validated by applying it to predict the future state condition of a low methane compressor based on given previous state data. Two cases have been studied using peak acceleration and envelope acceleration. The results show that the proposed method has potential to be a prediction tool for prognosis system based on time-series prediction.

2. Future Work

It has been stated in this dissertation that the demand of a reliable prognosis system is very important due to the potential advantages to be gained from reduced maintenance costs, improved productivity and increased machine availability. Even though the support vector predictor have demonstrated their ability in time-series forecasting schemes, advanced research needs to be done in several aspects before they can be applied to general real-time industrial application. These aspects are improving their application robustness (i.e., apply SMO solver instead of QP) to accommodate different system condition, mitigating the requirements for the representative data sets, improving the convergence properties, especially for complex operation applications. Schematically, the architecture of the forecasting tool is shown in Fig. 6.1.

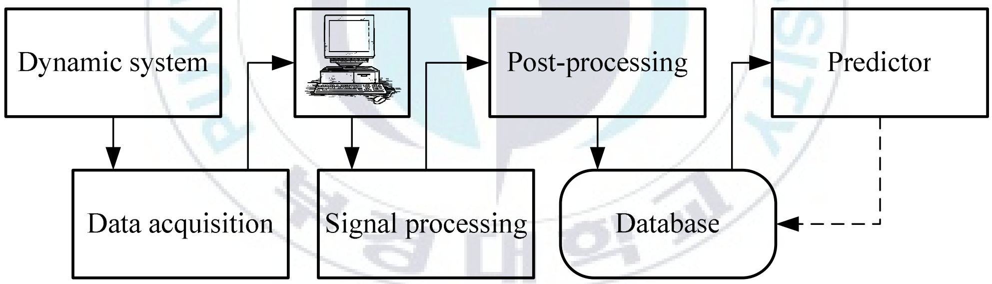


Fig. 6.1 The architecture of the prognosis based forecasting tool.

Signals are acquired from corresponding sensors, then after being properly filtered and sampled; the signals are transferred into computer. Signal processing is employed to generate the representative features from the acquired signals by applying different signal processing techniques. Then, post-processing is addressed to enhance the feature characteristic and derive the monitoring indices for forecasting operations.

기계 결함진단 및 예지를 위한 SVM

Achmad Widodo

부경대학교 대학원 기계공학부

국문 요약

설비 정비 관리 시스템(maintenance management system)의 한 핵심 분야인 기계 결함 진단(fault diagnosis) 및 예지(prognosis) 기술에 관한 연구는 관리 비용 저감, 생산성 및 기계 가용도 향상 등으로부터 얻어질 수 있는 무궁한 경제적 및 기술적인 잠재력으로 인하여 최근 전세계적으로 중요한 연구 과제가 되고 있다. 이와 관련된 연구들로 인공 신경망, 퍼지 전문가 시스템, 상태 기반 추론, random forest 등과 같은 지능 시스템에 기반을 둔 다양한 방법들이 개발되고 있다.

Support Vector Machine(SVM)은 종래 널리 사용되고 있는 인공 신경망 기법에 비해 탁월한 일반화(generalization) 능력을 가지고 있으므로, 높은 정확도를 가지고 기계 설비의 결함 진단을 위한 분류나 잔여 유용 수명의 예측을 할 수 있는 잠재 능력을 가지고 있다. 그러나 결함 진단 및 예지를 위한 SVM의 적용에 대한 연구는 세계적으로 아직 매우 드물다.

이 논문에서는 SVM 알고리듬을 기계의 결함 진단 및 예지를 수행하기 위한 지능 시스템으로서 확장하였다. SVM은 기계 학습의 프레임 워크에서 두 가지의 탁월한 능력인 분류와 회귀를 가진다. 결함 진단은 SVM의 분류 능력을 사용하여 수행되고, 기계 상태의 예지는 SVM을 이용한 회귀에 근거하여 수행된다. SVM은

주어진 데이터를 훈련할 수 있고 결과를 가중치로 저장하며, 분류 및 회귀를 수행하기 위하여 가중치를 사용한다. 원래, SVM은 선형 데이터의 두 클래스 분류 문제를 위하여 사용되었지만, kernel 사상(mapping)의 SVM에의 적용을 통하여 훈련 절차를 비선형 데이터를 이용한 훈련 절차 및 분류를 수행할 수 있다. 초평면(hyperplane)을 최적화함으로써, SVM은 분류 및 회귀 문제의 해결을 시도한다.

이 연구에서 제안된 방법은 결함 진단 목적의 분류를 위해 소위 특징 기반 기술에 기초하여 특징 추출 방법과 SVM을 조합한다. 특징 기반 기술은 센서로부터 취득된 진동, 전류 등의 다양한 신호의 원 데이터(raw data)를 통계치, 색깔, 모양 등과 같은 데이터가 가지는 다양한 특징으로 표현된다. 기계 결함 진단에서 특징은 기계 상태를 나타내는 값으로 방대한 종류의 특징들이 계산되므로 높은 차원의 데이터로 표현될 수 있다. 이러한 다량의 데이터는 전송 및 저장의 문제뿐만 아니라 분류 효율을 떨어트리는 문제가 발생할 수 있고 데이터의 차원을 저감할 필요가 있고, 이는 특징 추출 기법을 사용하여 해결할 수 있다. 특징 기반 분류 기술은 데이터 취득, 전처리(preprocessing), 특징 표현, 특징 계산, 특징 추출 및 선택 그리고 분류기(classifiers)로 구성된다.

제안된 방법은 SVM에서 분류 전략에 의하여 결함 진단을 수행하기 위하여 설계·제작된 유도 전동기 결함 실험 장치로부터 취득된 데이터를 사용하여 검증하였다. 결함으로는 굽은 축(bent shaft), 회전자 봉 결함, 질량 불평형, 상 불평형(phase unbalance) 및 편심 결함이 적용되었다. 전동기에 부착된 가속도계와 전류 프로브로부터 취득된 진동 가속도 및 전류 데이터가 사용되었다. 얻어진 진단 결과는 이 연구에서 제안된 방법이 결함 진단을 양

호하게 수행할 수 있고, 또한 향후 결함 진단 기술로 적용될 수 있음을 확인하였다.

예지는 열화 되고 있는 기계 요소 또는 하위 시스템(subsystem)의 잔여 유용 수명(residual useful life)을 정확하게 예측할 수 있는 능력으로 정의될 수 있다. 그러므로 신뢰할 수 있는 예측기(predictor)의 개발이 매우 중요하며, 그것은 동적 시스템의 임박한 상태를 예측하거나 회전 기계에서 손상 전파 경향을 예측하기 위하여 광범위한 산업에서 유용하게 적용될 수 있다. 예를 들어 기계 시스템에서 기계 성능 열화, 오작동 또는 파멸적인 고장을 예방하기 위하여 결함이 임계 값에 도달하기 전에 정확한 경보를 제공하기 위하여 예측된 정보는 상태 감시용으로 사용될 수 있다. 더구나 이는 제조 설비에서 수리 계획, 예지 정비 및 예방 정비의 계획 수립 그리고 예측 및 고장 허용 제어(fault-tolerant control)에 적용될 수 있다.

이 연구에서, SVM 기반 회귀 방법(SVM-based regression method)이 시계열 데이터의 예측기로서 사용될 수 있도록 확장되었다. 기계에서 취득된 경향 데이터(trend data)는 시계열로서 간주될 수 있고, 또한 운전 동안의 기계 정보를 포함한다. 제안된 방법은 이전 상태의 데이터에 기초하여 임박한 기계 상태를 예측하는데 사용된다. 제안된 방법을 검증하기 위해, 국내 석유 화학 플랜트에서 사용 중인 메탄 압축기에서 취득된 경향 데이터를 이용하였다. 예측기의 성능은 *RMS* 오차 및 상관 계수를 사용하여 평가되었고, 그 결과는 제안된 SVM 기반 회귀 방법이 신뢰할 수 있는 예지 도구로서 이용될 수 있는 잠재력을 가지고 있음을 확인하였다.