공학박사학위논문

대형 발전소 저널베어링 회전체 시스템의 비지도 학습 및 진단 기법 연구

Unsupervised Learning and Diagnosis Method for Journal Bearing System in a Large-scale Power Plant

2016년 2월

서울대학교 대학원

기계항공공학부

전 병 철

Abstract

Unsupervised Learning and Diagnosis Method for Journal Bearing System in a Large-scale Power Plant

Byung Chul Jeon
Department of Mechanical and Aerospace Engineering
The Graduate School
Seoul National University

Rotor systems are frequently used in machines and facilities for various industrial applications. Often the rotor systems fail to deliver their designed performance, thus resulting in substantial financial loss. These issues are very critical to some industrial sectors (such as power plants) where journal bearing rotor systems are regularly used. As a result, it is common to implement a diagnosis tool to such rotor systems. Automated diagnosis using data-driven techniques can enable detection of anomalies during early stages and can thus contribute to improved safety and increased cost savings. In the process of developing diagnosis algorithm, the robustness is one of the best important issues. Furthermore, for the diagnosis of a variety of fault conditions that may occur in a real system, the application of unsupervised learning techniques is needed.

In order to facilitate the development of robust diagnosis methodologies for the rotors in journal bearing systems, this research aims at advancing two essential research areas: Research Thrust 1 – datum unit optimization and Research Thrust 2 – omnidirectional regeneration (ODR) of gap sensor signals. In Research Thrust 1, the optimal datum unit will be defined by the comparison of separability and classification performance among feasible datum units. In Research Thrust 2, highly accurate and robust diagnosis approach using ODR signals will be introduced. The virtually generated ODR signals for circumferential direction can fully represent the vibration behavior of rotor system. Following the development of Research Thrust 1 and 2, Research Thrust 3 – unsupervised learning framework for power plant will give the basis of extension to the actual power plant diagnosis. Deep learning for unsupervised technique with high-level feature gives the reliable clustering results for power plant diagnosis.

Keywords: **Diagnosis**

Datum unit

Omnidirectional Regeneration (ODR)

Journal Bearing

Deep Learning

Student Number: 2012-30730

Contents

List of Tables

List of Figures

Nomenclatures

KLD = Kullback-Leibler divergence

FDR = Fisher discriminant ratio

PoS = probability of separation

DC = direct current

AC = alternating current

RMS = root mean square

ODR = omnidirectional regeneration

$s(f)$ = power spectral density function

$f(x)$ = probability density function

$F(x)$ = cumulative density function

$P(i)$ = probability mass function

μ_i = mean value of i^{th} class data

σ_i = standard deviation of i^{th} class data

Π = cost function value of genetic algorithm

PoS_j = PoS value of j^{th} feature

$\rho_{i,j}$ = correlation coefficient between i^{th} and j^{th} features

k = generation number of genetic algorithm

g_c^n = n^{th} feature of the feature set c

β = lower limit criteria of genetic algorithm

Chapter 1. Introduction

1.1 Background and Motivation

Rotor systems are frequently used in machines and facilities for various industrial applications. Often the rotor systems fail to deliver their designed performance, thus resulting in substantial financial loss. These issues are very critical to some industrial sectors (such as power plants) where journal bearing rotor systems are regularly used. As a result, it is common to implement a diagnosis tool to such rotor systems. Common fault types in rotating machinery include unbalance, misalignment, rubbing, cracking, and bearing failures [1, 2]. Although most of the faults in rotating systems can be identified by diagnostic specialists who visually inspect the spectral analysis of various signals, the need for an automated and reliable diagnosis system is steadily increasing. Automated diagnosis using data-driven techniques can enable detection of anomalies during early stages and can thus contribute to improved safety and increased cost savings. In the process of developing diagnosis algorithm, the robustness is one of the best important issues.

Automated diagnostic algorithms are typically developed using a supervised data from the test-bed or analysis model. However, for the diagnosis of a variety of fault conditions that may occur in a real system, the application of unsupervised learning techniques is needed. In recent years, deep learning is attracting attention as

unsupervised techniques. Deep architecture constructed by deep learning obtains the high level abstraction of given data, which can be used to generate features for machine learning. Deep learning outperforms in areas such as image recognition, speech recognition, handwritten character recognition. Deep learning can be applied to diagnosis of the actual power plant as unsupervised training using image data from vibration signals.

1.2 Overview and Significance

This research aims at advancing two research areas for increasing robust diagnostics of journal bearing rotor system: Research Thrust 1 – datum unit optimization for feature generation and Research Thrust 2 – Omnidirectional Regeneration (ODR) of gap sensor signals for diagnostic framework. Datum unit optimization can increase the robustness in the initial stage of diagnosis procedures whereas ODR influences over the entire diagnostic procedure. In order to develop automated diagnostic algorithms based on physics for various fault conditions, the data from test-bed and analysis model were utilized. In addition to advancing the diagnostic rules for the actual system, this research also makes the unsupervised diagnostics framework for a power plant, which will be elaborated in Research Thrust 3 – unsupervised learning framework for a power plant. Specifically, the research scope in this thesis is to develop technical advances in the following three research thrusts:

Research Thrust 1: Datum Unit Optimization

Research Thrust 1 addresses research challenges in data manipulation for feature extraction to improve the separatility. The performance of separation for feasible datum units in gap sensor signals are compared by three class separation metrics — Kullback-Leibler divergence (KLD), Fisher discriminant ratio (FDR), and a newly proposed measure: probability of separation (PoS). PoS are proposed to efficiently address the comparison of separability between various datum units by normalized, bounded and high sensitive manner. Furthermore, classification results are also introduced to verify that the selected datum units in separability comparison are well perform.

Research Thrust 2: Omnidirectional Regeneration (ODR) of Gap Sensor Signals

Research Thrust 2 suggests the generation of virtual sensing signals and its application to diagnosis algorithm. A journal bearing system generally uses two gap sensors installed in the right angle to obtain the vibration of the rotor. Employing the signals from two sensors at the fixed orientations may not detect direction-oriented anomalies like rubbing and misalignment. Omnidirectional Regeneration (ODR) of gab sensor signals can enable the extraction of virtual vibrations in any radial direction without additional installation of sensors. First the directionality of each

health state is quantified by the ODR signals. The diagnostics procedure for the directional anomalies is performed using ODR signals.

Research Thrust 3: Unsupervised Learning Framework for a Power Plant

Following the development of Research Thrust 1 and 2 for improvement of robustness of diagnosis rule, Research Trust 3 — unsupervised learning framework for a power plant will suggest a practical methodology for application to real system. Unsupervised learning is applicable to diagnosis of actual conditions out of known conditions which were analyzed by model or experimented. Recently, deep learning is considered as one of solution for clustering big unsupervised data. It is well known that deep learning outperforms in the recognition of big data for image, speech and handwritten character. After imaging the vibration signals of the various health conditions which may occur in the actual power plant, the deep learning diagnosis framework was proposed for unsupervised training.

The proposed optimal datum unit and ODR technique are expected to make significant contributions for improving robustness of journal bearing rotor diagnosis system. This advanced knowledge will be applicable to a broad diagnosis methodologies which uses gap sensor signals. It is also believed that imaging of gap sensor signals can be used simple high-level performance features which contains whole vibration characteristics. The improved robustness of diagnostic procedures

and deep learning technique with vibration image gained from this research will facilitate the advances of diagnostics methodologies for the actual power plant.

1.3 Thesis Layout

The thesis is organized as follows. Chapter 2 reviews the current state of knowledge related to data-driven diagnosis, diagnosis of rotors in a power plant, and deep learning for unsupervised training. Chapter 3 describes the journal bearing rotor system and presents the works related to diagnostic module. Chapter 4 presents the methodology for datum unit optimization for robustness of feature extraction. Chapter 5 presents Omnidirectional Regeneration (ODR) of gap sensor signals to employ suitable diagnosis technique for directional health conditions. Chapter 6 represents the unsupervised learning framework using deep learning for the actual power plant diagnosis. Finally, Chapter 7 summarizes the contribution of the research work and suggests the insight on future works.

Chapter 2. Literature Review

In this chapter, the state-of-art knowledge for data-driven diagnosis within the scope of this thesis will be reviewed: (1) data-driven diagnosis approach, (2) diagnosis of rotor system in a power plant, and (3) deep learning for unsupervised training.

2.1 Data-driven Diagnosis Approach

Fault diagnosis for mechanical systems has been attracting considerable attention from researchers in many application areas. With the development of artificial intelligence techniques and the Improvement of computational abilities, data-driven fault diagnosis methods have advanced gradually in the past decades. A commonly used data-driven diagnosis technique [3-8] consists of four parts; data acquisition, feature extraction, feature selection, and classification prediction.

Data Acquisition

Data acquisition describes the step of gathering and sorting useful signals from targeted engineering system by using appropriate sensors for the purpose of fault diagnosis. This process is the fundamental step in implementation of the fault diagnostic processes. Data collected in a condition-based maintenance program can be categorized into two main types: event data and condition monitoring data. Event

data include the information on what happened (e.g., installation, breakdown, overhaul, etc., and what the causes were) and/or what was done (e.g., minor repair, preventive maintenance, oil change, etc.) to the targeted physical asset [9]. Condition monitoring data are the measurements related to the health condition of the physical system [9]. Condition monitoring data can be measured in a variety of forms such as vibration, temperature, pressure, humidity, acoustic, etc. Among many types of signals, vibration data are the most widely used for the diagnosis of a rotating system because they directly represent the condition of system [10, 11]. Vibrations can be measured as three forms of signals: displacement, velocity, and acceleration. The vibration measure among the three forms is selected based on the frequency content of vibration present, the type of machine, and the type of the analysis to be conducted [12]. For low frequencies (<100 Hz) displacement measurements are appropriate; for mid frequencies (50-2000 Hz) velocity measurements are appropriate; and for high frequencies (>2000 Hz) acceleration measurements are appropriate [13]. In a journal bearing system of a power plant, two gap sensors in orthogonal axes are installed in the consideration of the targeting frequency and the type of system.

Feature Extraction

Feature extraction is the process to define a mapping from the acquired signals to useful information which represents more easily separable. Various signal processing methods for vibration data have been developed, such as angular resampling [14-16], statistical method [17], Wigner-Ville distribution [18, 19],

principle component analysis [20, 21], independent component analysis [22], empirical mode decomposition [23, 24], and wavelet transform [25-27]. Among these techniques, angular resampling is the most popular technique for journal bearing systems which have no contact between mechanical components in normal conditions [28]. The angular resampling process regenerates vibration signals based on the rotation angles and facilitates robust diagnosis of a journal bearing rotor system. Even under steady-state operating conditions, resampling can reduce the randomness of the data that originates from uncertain conditions like rpm, temperature, manufacturing tolerance, etc. Time- and frequency-domain features can be generated using angular resampled vibration signals. As signal features for data-driven fault diagnosis of rotor systems, statistics-based features, such as root mean square (RMS) and kurtosis, are commonly used in the time-domain, while features related to fundamental frequency and its harmonics are widely used in the frequency-domain [29, 30].

Feature Selection

Feature subset selection is an effective way for reducing dimensionality, eliminating irrelevant data and redundant data, and increasing accuracy [31]. It can reduce the computational complexity as well as improve the classifier's generalization ability [32]. Although two features may carry good classification information when treated separately, there is little gain if they are combined into a feature vector because of a high mutual correlation. Thus, when the highly correlated

features are used in the same time, complexity increases without much gain of separation performance. Additionally, fewer features require less run time to train and to apply the classifier. Increasing the generalization ability can be considered as reducing the risk of overfitting. The higher the ratio of the number of training samples N to the number of feature dimensions, the better the generalization properties of the resulting classifier [32]. Feature selection is the important issues in pattern classification for machine diagnosis. It has three goals: reducing the cost of extracting features, improving the classification accuracy, and improving the reliability of the estimate of performance [12]. The simplest algorithm for feature subset selection is to test each possible subset of features finding the one which minimizes the error rate [31]. This is an exhaustive search of the space, and is computationally intractable for all but the smallest of feature sets. Usually, feature selection methods can be divided into three categories: exponential algorithms, sequential algorithms and randomized algorithms. Exponential algorithms contain exhaustive search, brand and bound, and beam search. These algorithms can find the best feature subsets whereas the calculation time is very expensive. Sequential search takes feature dependencies into account, which characteristic is fast speed whereas the best optimal solution usually cannot be obtained. Genetic algorithm (GA) is one of the randomized approach which can find feature subsets in good performance with low computation complexity. GA can deal with large search spaces efficiently, and hence has less chance to get local optimal solution than other algorithms [33].

Classification Prediction

In a supervised learning, each state of training data is labeled as an individual class. Classification algorithm identifies the class of the testing data by a trained classifier. Training features extracted from training data are used to define statistical characteristics for learning classifiers. By using the trained statistical rules, the testing features extracted from testing data are classified into one of the trained classes. There are many statistical techniques can be used to design a classifier. Principal Component Analysis (PCA) which calculate the directions (components) of data according to the projection by finding eigenvectors maximizing variance is used to reduce the dimension of data and classify large number of data [34-36]. Also a statistical method, Fisher discriminant analysis (FDA) was originally widely used in pattern classification [37]. FDA seeks directions that are efficient for class discrimination whereas PCA finds directions that are only efficient for representation of whole data [38]. Linear statistical approach like PCA and LDA can represent high performance in a large number of high dimensional data processing. However, FDA has a limitation which may give distorted results for significantly non-Gaussian data. Artificial neural network (ANN) based on the biological learning process of the human brain is used very extensive in pattern recognition [39, 40]. ANN has advantages which are requiring less formal statistical training, ability to implicitly detect complex nonlinear relationships, ability to detect all possible interactions between predictor variables [41, 42]. However, it has also disadvantages which

include its “black box” nature, great computational burden, proneness to overfitting [41, 43]. Support vector machine (SVMs) is based on statistical learning theory and they specialize for a smaller sample number because it minimize structural risk of given data [39, 44]. SVM finds the hyperplane which make the maximum-margin distance between the closest data points of each class to the hyperplane. The larger the margin of SVM contributes the lower the generalization error of the classification prediction. PCA, FDA and SVM was suggested as linear classifier initially. However, they can be used for nonlinear pattern recognition problem by using Kernel method [45, 46].

Summary and Discussion

As discussed above, generic procedures of data-driven diagnosis approach for engineering systems have been well established. Especially a wide variety of signal processing techniques have been developed for efficient and effective feature extraction. On the other hand, there have been only a few studies related to define datum although the unit of datum may affect the performance of reasoning the problems. Prior work has shown that the sampling rate of data acquisition and the amount of data available for diagnosis have an effect on the performance of the diagnosis algorithm.[47, 48] This implies that an optimum datum unit must be decided with much care for robust fault diagnosis of rotor systems. Thus, research is needed to determine the optimum datum unit for the given system.

2.2 Diagnosis of Rotor System in a Power Plant

Rotor systems in a power plant which are commonly supported by journal bearings are one of the most critical components. Major fault types of rotors in a power plant include unbalance, misalignment, rubbing, cracking, and bearing failures [1, 2]. Typical malfunctions which have their representations in the frequency range and their corresponding symptoms have been summarized in Table 1 [10, 49-51]. For the diagnosis of rotor system, vibration is considered as the most important symptoms of system behavior because it has high content of information, non-intrusive and easy applicability, and well-developed data processing methods [50]. The Depending on the vibrational signal acquisition and processing methods rotor, diagnostic techniques are divided in many ways.

Table 2-1 Typical steam turbine malfunctions and their vibrational symptoms

Malfunction	Typical symptoms
Unbalance	1x f_0 component in vertical and horizontal directions, constant amplitude and phase, decreasing at low rotational speed
Misalignment	2 x f_0 component in vertical and horizontal directions, 'banana-shaped' or flattened shaft orbits, high harmonic components in axial direction
Rubbing	Super-harmonics or sub-harmonics, 'Truncated' signal for rubbing zone
Oil whirl and whip	0.42~0.48 x f_0 component in vertical and horizontal direction

Permanent rotor bow	1 x f_0 component in vertical and horizontal direction (also at low rotational speed), strong correlation between 1 x f_0 components in vertical and axial directions
Rotor crack	Continuous changes of 1 x f_0 and 2 x f_0 components amplitudes and phases during steady-state operation, reduction of critical speeds and increase of vibration amplitudes on passing through them
Bearing problems	Increase of sub-harmonic components (typically slightly below 0.5 x f_0), relative vibration increase, shaft orbits with loops, high and unstable amplitudes of higher harmonic components, sensitive to bearing oil pressure

Faults and its diagnosis methods

All rotating machinery has an essential degree of unbalance which is one of the most common fault in rotordynamics. Therefore it is considered as a fault when the magnitude of vibration exceeded the limit level for the system [52]. Recently Ganeriwala et al. [53] suggested detecting unbalance by using operating deflection shapes (ODS). These studies showed the identification of unbalance by estimating differences of ODS between baseline and seven unbalance cases. As a physics-based approach for unbalance fault, Sudhakar and Sekhar [54] proposed the methodology of fault identification for unbalance using theoretical simulation model. They used an equivalent load minimization method which characterizes unbalance faults by minimizing difference between equivalent loads estimated in the system due to the fault and theoretical fault model. The assessment of unbalance location can considered as a unique part of challenges. Some recent works deal with localization

of unbalance of rotor system. Yang and Hsu [55] proposed orderly searching the possible location of faults instead of searching all combinations, and Walker et al. [56] have been localized unbalance through data driven approach using the artificial neural networks.

Misalignment is the second most common malfunction type in a rotor system [57]. 2X component in a large steam turbines is usually associated with misalignment, which has been confirmed both theoretically and experimentally [58]. frequency in Redmond introduced the theoretical study of misalignment in coupled rotating machinery [59]. The complex system forces and motions are derived by application of the Lagrange method without the imposition of specific harmonic-excitation assumptions. Additionally, a lot of the misalignment model studies have been conducted to demonstrate the physics and understand physical meaning [57, 59-64]. Tejas and Ashish showed the vibration response of misaligned rotors and its diagnostic results [60]. The coupled rotor system was modeled using Timoshenko beam elements with all six degree of freedom. To extract some diagnostic features, full spectra and orbit plots are effectively used to reveal the unique nature of misalignment fault leading to reliable misalignment diagnostic information. Sarkar et al. proposes a method for computing the displacement-dependent stiffness terms from the experimental static load-displacement data [57]. The orbit of the rotor around the static equilibrium is determined using a time-integration scheme.

Rub is commonly caused as a secondary fault from unbalance or misalignment and can lead the catastrophic failure of rotating system abruptly. The mechanical contact between stationary parts of machine and rotating elements can generate a variety of rub conditions. There exists some experimental studies for the purpose of rubbing behavior demonstration and development of diagnostic algorithms [65-70]. Xiaozhang demonstrated the impact and rub phenomenon of a rotating rotor with a fixed boundary [65]. Wang and Chu studied the determination of the rubbing location by using wavelet transform for the acoustic signals [68]. Dai et al. showed the numerical simulation of partial and full rubbings using simple Coulomb friction model and the multiple segments linear spring model, then experiments are demonstrated to verify the feasibility and the limitation of using stops to control the large vibration of rotors [69]. Chu and Lu established an experimental setup of a rub-impact multi-disk rotor system which can demonstrate the condition of the full rub and several severity level of impact rubbing [70]. The authors showed the result of vibration due to a certain level of partial rubbing and full rubbing in the form of orbit plot and the existence of a certain frequency components such as 1/2X, 3/2X, 2X, 3X, etc.

Shaft crack is potentially catastrophic fault in rotating machinery, and so early detection is very important. Cracked shaft may decrease the critical speed of the shaft and also exhibit a nonlinear vibration behavior leading to increases the amplitude of 1X and 2X harmonics [71]. Numerous researchers have used theoretical methods for

the research of shaft crack, whereas, others have employed both theoretical and experimental ones, and a few have focused on only experimental methods. Huang et al. studied the dynamic response of a rotor model with a shaft crack, by introducing the governing equations in numerically [72]. The authors examined the frequency response of the shaft model with different crack depths and locations, and showed that the 1X and 2X harmonics were excited. Luo et al. set up for the rotor-bearing model with coupling faults of crack and rub-impact [73]. The results presented that there are unique dynamic characteristics of rotor bearing system with cracks coupled with rub-impact fault which are different from the ones with only one type of fault. Varney et al. introduced analytical model and designed an overhung rotor test rig to monitor seal face dynamics [74]. Experimental results of angular response orbits and the magnitude of the 2X harmonic tilt response are provided for cracks between 0% and 40% depth.

Fluid-induced instability (so called oil whirl and whip) is one of the most common cause of subsynchronous faults in a journal bearing system. It can result in secondary anomalies such as fatigue, wear, and extensive damage to rotor systems. These fluid-induced instabilities can be occurred in oil lubricated bearings, bearing seals, and clearances between blade and housing. Several studies on the fluid instability of the rotor system using modeling and simulation were carried out in the last few years. De Castro et al. suggested nonlinear mathematical models for a rotor-bearing system to study the fluid-induced instability [75]. The models are used to predict oil whirl

and whip instabilities in different unbalance conditions. The results of model simulations were compared with the data at a vertical power plant and a horizontal test-bed. The authors claimed that nonlinear solution to simulate hydrodynamic journal bearing system shows sufficiently accurate vibration behavior for predicting instabilities. Fan et al. studied to predict fluid-induced instability by the experimental setup and developed signal processing methodologies [76]. The authors suggested a Hilbert spectrum combining a full spectrum, then the transient position of a shaft centerline combining an acceptance region is investigated to predict instability at an early stage.

Artificial Intelligence Approach

Artificial intelligent techniques, such as expert systems, fuzzy logic, and artificial neural networks (ANNs), support vector machine (SVM), etc., have been successfully applied to automated detection and diagnosis of machine conditions [77]. The expert systems [78, 79] with the information from system knowledge were developed to diagnosis the fault of steam turbine generator sets (STGS). However, it is difficult to design an efficient reasoning methods which can lead appropriate results from a large amount of rule-based knowledge. Besides the expert systems are sometimes limited for real-time condition monitoring of STGS since it requires sufficient time for the inference process.

Fuzzy reasoning [80-82] is expressed as a form of many-valued logic and has been proposed to diagnosis vibration anomalies of rotating machinery using inaccurate and uncertain information. Similarly to the expert systems, fuzzy reasoning systems also depend heavily on the user's experience for the construction of the fuzzy sets and their associated membership functions. For designing fuzzy rules in high performance, it requires the expert's knowledge or trial and error database for possible conditions. It needs expensive cost for development and maintenance. Therefore, the fuzzy reasoning may not be applicable for the fault diagnosis of practical rotating system using vibration signals.

Artificial neural network (ANN) [83-86] determines optimized relationships between input and output nodes for reasoning or regression problems. It was suggested as a real-time response operator in practical diagnostics applications. The ANN using error back propagation can train the nonlinear relationships from supervised data, then introduce the desired reasoning results from testing data for system diagnosis. However, it has still some drawbacks such as slower learning speed and the convergence in local minimum. In order to apply anomaly diagnosis of practical rotor systems using vibration signals, the performance and efficiency must be further tested.

Support vector machines (SVMs) are suggested as an intelligent fault diagnostic method for rotating machinery [87]. SVM is a supervised learning method which determines optimized hyper-planes by solving the problem of structural risk

minimization. It is widely used for classification and regression analysis using linear or nonlinear data [45, 88, 89]. Some researches were suggested that SVM is more efficient than ANN methods [90, 91] because the difference of a risk minimization function. SVM uses the structural risk minimization which has high performance in generalization, while ANN minimizes the traditional empirical risk.

Summary and Discussion

There have been many studies for major faults of a rotor system in a power plant and its diagnosis methodology. Most of diagnosis techniques with the consideration of physical meaning deals the target faults in one or two. Then sometimes it gives the qualitative diagnostic results which need the additional judgement in engineering sense. In order to overcome these problems, AI approach for autonomous diagnosis such as expert system, fuzzy reasoning, ANN, SVM, etc. was suggested. However, the expert system and fuzzy reasoning need heavy prior information to construct the algorithm as well as has heavy dependence of operator's experience. The other AIs like ANN or SVM are not considered the physical meaning of the system. The understanding of system behavior and characteristics of gathered signal is very important part in the development of diagnostic system. Gap sensors which are mounted generally two sensors in fixed orientation may indicate different signal for the similar anomaly states when the occurrence direction of the fault. Therefore, it is necessary to develop the diagnosis algorithm with the consideration of the characteristics of gap sensor signal.

2.3 Deep Learning for Unsupervised Training

In the region of artificial intelligence, deep learning has been received high attention over the past decade. Deep learning is a branch of machine learning techniques to model high-level feature abstractions in data by using multiple layers of processing [92-94]. It has shown remarkable performance in image recognition [95-97], speech recognition [98, 99], and handwritten character recognition [100-102]. The major reasons for the popularity of deep learning these days are the significantly increased computational abilities (e.g., general-purpose graphical processing units or GPGPUs), the highly increased size of training data, and the advances in signal processing and machine learning techniques.

Deep belief network (DBN) is well-known deep learning algorithm, which is hierarchical multilayer neural network models to learn high-level representation for their input data [103]. DBN consists of several stacked restricted Boltzmann machines (RBMs) which is energy based and unsupervised training methodology.

Restricted Boltzmann Machine (RBM)

RBM is a two layer network in which stochastic, binary pixels are connected to stochastic, binary feature detectors using symmetrically weighted connections [104]. The simplest type of RBM is shown in Figure 2-1. The input data correspond to visible units **v** because their states comes from observations; the feature detectors correspond to hidden units **h.** Weights are connected between visible and hidden

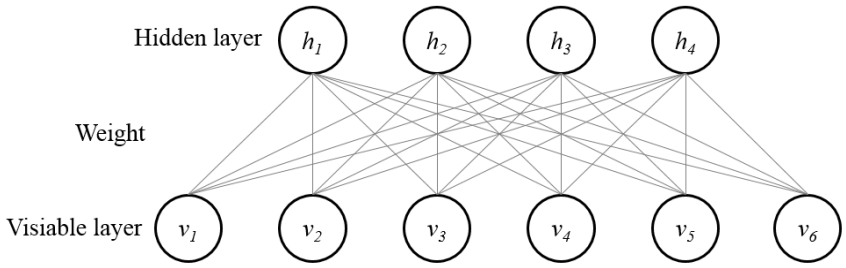


Figure 2-1 Structure of a restricted Boltzmann machine

units symmetrically while it is restricted the connections between visible-visible or hidden-hidden units. These restriction of inter-layer connection in RBM makes efficient and suitable building blocks for training DBNs.

A joint configuration, (v, h) of the visible and hidden units has an energy[105] given by:

$$E(\mathbf{v},\mathbf{h};\theta) = -\mathbf{v}^{\mathrm{T}}\mathbf{W}\mathbf{h} - \mathbf{b}^{\mathbf{T}}\mathbf{v} - \mathbf{a}^{\mathbf{T}}\mathbf{h}$$
$$= -\sum_{i=1}^{n}\sum_{j=1}^{m} W_{ij}v_i h_j - \sum_{i=1}^{n} b_i v_i - \sum_{j=1}^{m} a_j h_j \qquad (2.1)$$

where $\theta = \{\mathbf{W}, \mathbf{b}, \mathbf{a}\}$ are the model parameters, W_{ij} is the symmetric interaction weight between visible unit i and hidden unit j; bi and aj represent biases of visible and hidden units; v_i and h_j are the binary states of visible unit i and hidden unit j. The energy based joint distribution over the visible and hidden units is defined by:

$$P(\mathbf{v}, \mathbf{h}; \theta) = \frac{1}{Z(\theta)} \exp(-E(\mathbf{v}, \mathbf{h}; \theta)) \tag{2.2}$$

$$Z(\theta) = \sum_v \sum_h \exp(-E(\mathbf{v}, \mathbf{h}; \theta)) \tag{2.3}$$

where Z(θ) represents the partition function. The probability of the visible units is assigned as:

$$P(\mathbf{v}; \theta) = \frac{1}{Z(\theta)} \sum_h \exp(-E(\mathbf{v}, \mathbf{h}; \theta)) \tag{2.4}$$

On account of the special bipartite structure of RBM, the hidden units can be definitely marginalized as:

$$P(\mathbf{v}; \theta) = \frac{1}{Z(\theta)} \sum_h \exp(\mathbf{v}^{\mathrm{T}}\mathbf{W}\mathbf{h} + \mathbf{b}^{\mathbf{T}}\mathbf{v} + \mathbf{a}^{\mathbf{T}}\mathbf{h}) \tag{2.5}$$

The conditional probability over visible vector v and hidden units h can be derived from equation (2.2) and are activated by logistic functions:

$$P(\mathbf{v}|\mathbf{h}; \theta) = \prod_{i=1}^{n} p(v_i|\mathbf{h}) \tag{2.6}$$

$$P(\mathbf{h}|\mathbf{v}; \theta) = \prod_{j=1}^{m} p(h_j|\mathbf{v}) \tag{2.7}$$

$$p(v_i = 1|\mathbf{h}) = \delta(b_i + \sum_{j=1}^{m} W_{ij} h_j) \tag{2.8}$$

$$p(h_j = 1|\mathbf{v}) = \delta(a_j + \sum_{i=1}^{n} W_{ij} v_i) \tag{2.9}$$

where $\delta(x) = 1$ (1+ exp(−x)) is the logistic function used as an activation function. The derivative of the log-likelihood with respect to the model parameters θ can be obtained from equation (2.4):

$$\frac{\partial \log P(\mathbf{v};\theta)}{\partial W} = E_{P_{data}} < \mathbf{v}\mathbf{h}^{\mathrm{T}} > - E_{P_{model}} < \mathbf{v}\mathbf{h}^{\mathrm{T}} > \tag{2.10}$$

$$\frac{\partial \log P(\mathbf{v};\theta)}{\partial a} = E_{P_{data}} < \mathbf{h} > - E_{P_{model}} < \mathbf{h} > \tag{2.11}$$

$$\frac{\partial \log P(\mathbf{v};\theta)}{\partial b} = E_{P_{data}} < \mathbf{v} > - E_{P_{model}} < \mathbf{v} > \tag{2.12}$$

where $E_{P_{data}} <\cdot>$ represents an expectation with respect to the data distribution $P_{data}(\mathbf{h}, \mathbf{v}; \theta) = P(\mathbf{h}|\mathbf{v}; \theta) P_{data}(\mathbf{v})$; $P_{data}(\mathbf{v}) = \frac{1}{N} \sum_k \delta(\mathbf{v} - \mathbf{v}_k)$ is the empirical distribution; $E_{P_{model}} <\cdot>$ represents an expectation with respect to the distribution defined by the model. Learning process in RBM is achieved by a very simple learning rule for approximation to the gradient of a different objective function, called the "Contrastive Divergence" (CD) [106]. The equation of approximated weight function is represented as:

$$\Delta W_{ij} = \alpha\left(E_{P_{data}} < \mathbf{vh}^{\mathrm{T}} > - E_{P_T} < \mathbf{vh}^{\mathrm{T}} >\right) \tag{2.13}$$

where α represents the learning rate and P_T is a probability distribution defined by running an alternating Gibbs Markov chain, initialized by setting the binary states of the visible units to be the same as a data-vector, for T full steps. Setting T = q (or CD_q) represents the indication of learning using q full steps of alternating Gibbs sampling. In the process of RBM training, the CD learning with T = 1 (or CD_1) has been shown efficient and high performance [107]. Nevertheless RBM may learn better in the case of more steps (e.g. T = 10 or above) of alternating Gibbs sampling [108].

Deep Belief Networks (DBN)

DBN is a probabilistic generative models, containing multiple layers of hidden units, which can be defined as a composition of unsupervised training networks such as RBMs. The DBN structures are shown in Figure 2-2. Firstly, a RBM is used to learn the parameters between observation and hidden layer using unsupervised manner of training. After learning the weighting parameters, the hidden units of first RBM are used as visible units of another RBMs in order to generate higher level feature representation. These step-wise training technique is fast and effective [94]. After training of stacked RBMs, supervised fine-tunings are performed in DBNs.

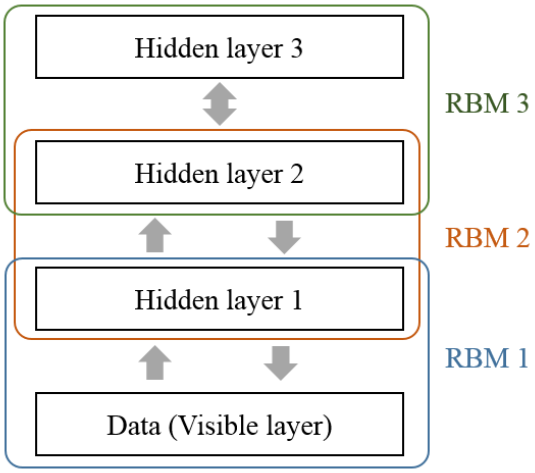


Figure 2-2 Structure of a deep belief network

The initial parameters are sequentially retrained by back-propagation from the DBN results to initial observations [109, 110].

Summary and Discussion

The high performance of feature abstraction representing the characteristics of given data have enabled the deep learning methods to effectively exploit complex or nonlinear reasoning problems. Deep learning algorithms in superior generalization performance may be applied for supervised classification and unsupervised pattern analysis using high-level feature representation [92]. This learning approach can give diagnostic efficiency for the diagnosis module of the actual power plant using vibration signals because it already greatly outperformed in areas such as image recognition, speech recognition, handwritten character recognition. However, there

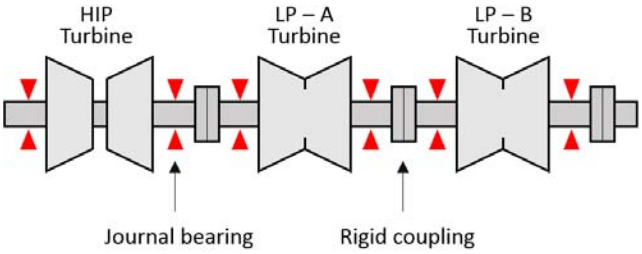
is a need for specific research about diagnostic framework and the transformation of information from vibration data.

Chapter 3. Journal Bearing Rotor System and Diagnostic Module

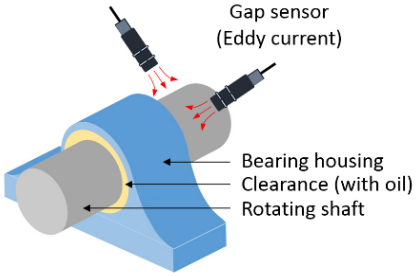
3.1 Overview of the Journal Bearing Rotor System and Its Behavior

3.1.1 Journal Bearing Rotor System Used in a Power Plant

As Journal bearings are frequently used in systems that operate in high-load and high-speed conditions. These operating conditions lead to a high risk of system malfunction. Power plant turbines are generally composed of three stages, including a high and intermediate pressure turbine and two low-pressure turbines, as shown in Figure 3-1(a). Each stage of turbine is supported by journal bearings, as illustrated in Figure 3-1(b). Journal bearings are composed of a fixed housing, a rotary shaft, and an oil film between the shaft and housing. Oil films support the journals while the shaft rotates. Unlike ball or roller bearings where accelerometers are widely used for acquisition of vibration signals, a gap sensor is used in journal bearings to acquire vibration signals in a displacement form. A gap sensor measures the radial position of the shaft by the variation of eddy current. Through two sensors positioned in a right angle, an orbit plot representing the motion of the shaft centerline can be generated. At the same time, the average shaft centerline position can be measured from the DC component of the gap sensor signal. The AC component of the gap



(a)



(b)

Figure 3-1 Schematics for (a) rotor system of a power plant (b) journal bearing

sensor signal represents the relative vibration of each revolution. These data give information on the behavior of the overall shaft motion in journal bearing systems.

3.1.2 Test-bed for Rotor in a Journal Bearing System

An RK4 journal bearing test-bed was used to demonstrate how effectively the optimum datum units diagnose potential faults in a journal bearing rotor system. This section introduces the fundamentals of the test setup and describes how anomaly conditions were seeded in the rotor system for testing.

This study is based on the data collected from an RK4 journal bearing test-bed, a product of GE Bently Nevada. As shown in Figure 3-3, a long shaft with an 800g disc attached is connected by a coupling to a short shaft. The whole system is

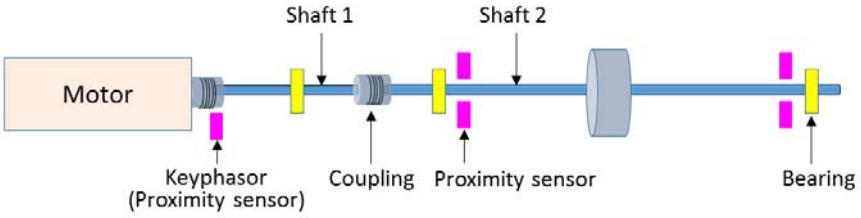
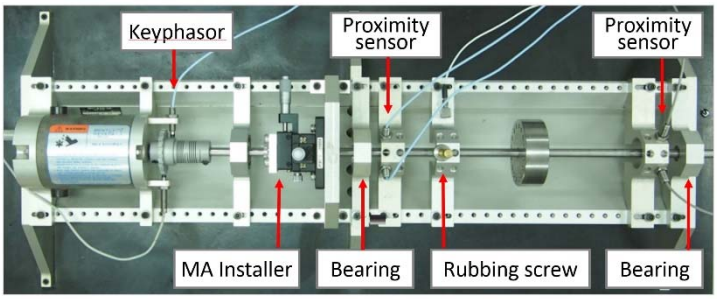
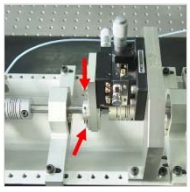


Figure 3-2 Schematic diagram of an RK4 test-bed

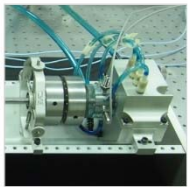


(a)

(b)



(c)



(d)

Figure 3-3 (a) RK4 test-bed and setups for (b) rubbing (c) misalignment, and (d) oil whirl

supported by three journal bearings. At the speed of 3,600 rpm, four health states were implemented into the journal bearing test-bed: normal, rubbing, misalignment, and oil whirl. The gap signals were measured at two locations on the shaft adjacent to the journal bearing and two proximity sensors were mounted at right angles in each location. One channel of a keyphasor was also located right next to the motor as a reference signal for each rotation.

A bit of imbalance generally exists in the rotor system even under the normal condition. The normal condition is set with an RMS level of 20 μm, which is equivalent to 40% of ISO 7019-2, that is, a vibration regulation for normal condition of a steam turbine in excess of 50MW at 3,600 rpm [111]. Each of the three anomaly conditions—rubbing, misalignment, and oil whirl—was individually implemented to the rotor system under the normal condition. A rubbing condition was created by directly rubbing a screw on Shaft 2. This condition is precisely named a *point impact rubbing*. To control the level of the rubbing condition consistently, an acceleration sensor was attached to the screw jig and the data within the acceleration of 2.0±0.3 m/s^2 were considered those obtained under the rubbing condition. A misalignment condition was made by shifting the end of Shaft 1 horizontally 20 μm to the right using a misalignment jig. This condition is called an angular misalignment. The oil whirl condition was induced using an oil whirl toolkit that is designed by the RK4 manufacturer. The condition was maintained by controlling the oil pressure at 1.8±0.2 psi.

Four channels of the gap signal and one channel of the keyphasor signal were acquired via five proximity sensors in the RK4 test-bed. This study used a 3300 proximity transducer, produced by GE Bently Nevada. The gap sensor measures the position of the rotating shaft by obtaining the eddy currents due to the interaction of the rotating shaft and the sensor. The signal of a proximity sensor is obtained as a voltage form comprised of direct current (DC) and alternating current (AC). The DC component provides the absolute position of the rotating shaft. The AC component represents the relative vibration of each revolution. These data give information of the overall shaft motion in the journal bearing system. The gap signals from the proximity sensors were acquired at 8.5 kHz via NI DAQ 4432. Every test to emulate the four health states—one normal and three anomalies—was conducted for sixty seconds and the test for each health condition was repeated twice. The acquired data sets were used to determine the optimal datum units by evaluating the capability of class separation.

Figure 3-4 gives the graphical interpretations of the four health states in the time- and frequency-domains. The first-column plots represent the waveforms of gap signals acquired from four health states; the second column displays the orbits; the third exhibits the frequency responses of the resampled data. Each health condition has a different waveform and frequency response. For the normal case, a basic sinusoidal wave was confirmed with the time plot, and the frequency response indicated that the 1x frequency is dominant. In the rubbing scenario, the effect of

truncation from the rubbing screw was well observed in the wave. As the result, the frequency response also showed a slight increase in harmonics compared to that of the normal case. For the misalignment case, the heavy preload due to misalignment contributed to the figure-eight shaped orbit. This is expressed with the observed high

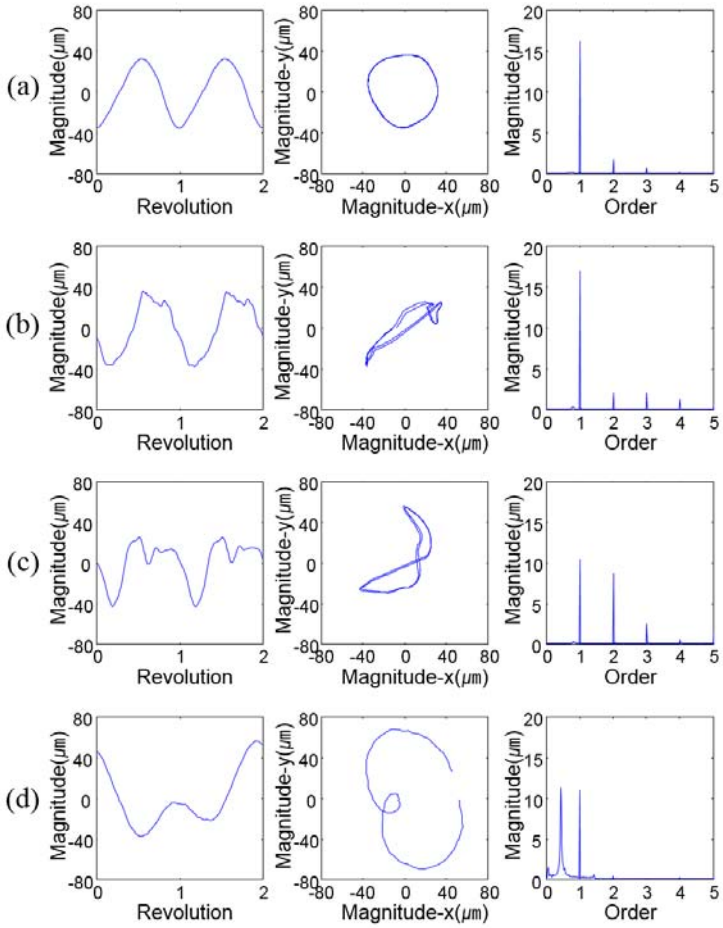


Figure 3-4 Time, orbit, and frequency response plots of RK4 for (a) normal (b) rubbing (c) misalignment (d) oil whirl experiment results

2x component of the frequency response. For the oil whirl scenario, the 0.42-0.48 times of the fundamental frequency become apparent.[10, 49] Indeed, the 0.45x component became highlighted in the RK4 data; this was also perceived in the waveform. In some situations of oil whirl, other sub-harmonic components (i.e., 0.43x) can be dominant.

3.1.3 Physics of rotor system in journal bearing using FEA

FEA model corresponding to the test-bed was employed for double checking the physics of rotor in a journal bearing system. FEA model also has some advantages in the aspect of time and cost saving compared to the experiment. The analysis model enables the examination of system behavior and the gathering additional vibration signals for various levels for anomaly.

The general dynamic equation is:

$$\mathbf{M}\ddot{\mathbf{x}}(t) + (\mathbf{C})\dot{\mathbf{x}}(t) + (\mathbf{K})\mathbf{x}(t) = \mathbf{F}(t) \tag{3.1}$$

where **M, C** and **K** are matrices of mass, damping and stiffness, and **F** is the external force vector. In rotordynamis, Equation (3.1) gets additional terms from gyroscopic effect **G**, and the rotating damping effect B leading [112]:

$$\mathbf{M}\ddot{\mathbf{x}}(t) + (\mathbf{C} + \mathbf{G})\dot{\mathbf{x}}(t) + (\mathbf{K} + \mathbf{B})\mathbf{x}(t) = \mathbf{F}(t) \tag{3.2}$$

Analytical model of the rotating system was developed by using the ANSYS Classic. The elements used are as follows: 3-D linear finite strain beam element (beam188) for shaft, 2-D spring element (COMBI214) for bearing, spring-damper (COMBIN14) and lumped mass combined element for flexible coupling. Material properties of analysis model were adjusted by the values measured from the test-bed. The stiffness of model was calibrated by the natural frequency of test-bed through impact hammer test. The natural frequencies of experiment and analysis are listed in Table 3-1. Modeling of flexible coupling used equivalent symmetric spring and lumped mass combined element because matrices of gyroscopic and rotational damping are supported to the only symmetric geometry in ANSYS. Then translational and rotational stiffness are calibrated using the natural frequency of system.

Table 3-1 Natural frequencies of experiment and analysis model

Model		Experiment		Analysis	
		free-end	on test-bed	free-end	on test-bed
Natural frequency [Hz]	1st bending	18.5	44.5	18.2 18.5	46.4
	2nd bending	128.0	-	136.5 137.3	82.8
	3rd bending	360.5	-	345.5 346.4	331.9

FEA model analyzed the health states of normal and rubbing cases. All rotating machinery in normal condition has an inherent degree of unbalance. The force due to unbalance in rotors are defied as:

$$F_{unb} = m \cdot r \cdot \omega^2 \tag{3.3}$$

where m, r, and ω are mass of rotor, distance between unbalance and the axis of rotation and angular frequency. The unbalance force are implemented to model by F_x and F_y in Equations (3.4) and (3.5).

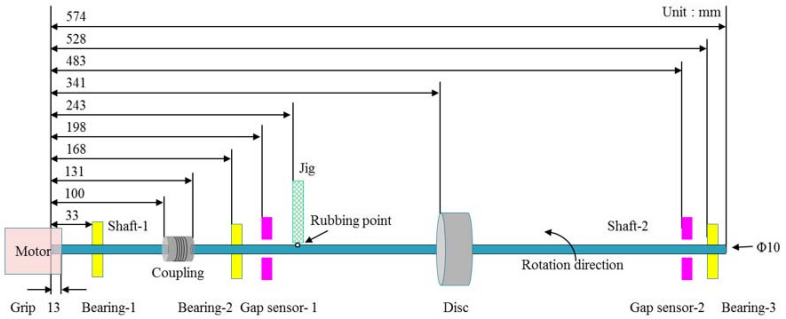
$$F_x = F_{unb} \cdot \cos(\omega(t)) \tag{3.4}$$

$$F_y = F_{unb} \cdot \sin(\omega(t)) \tag{3.5}$$

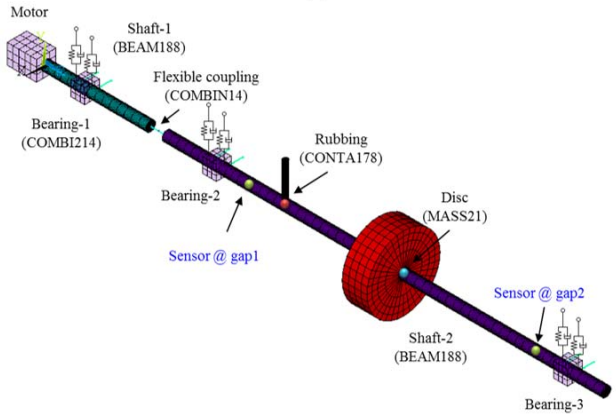
Rubbing conditions are modeled using CONTA178 element which executes node to node contact for the rubbing zone [113]. The penalty stiffness was employed between shaft and contact element.

$$\mathrm{F}_{normal} = k_{normal} \cdot x_{penentration} \tag{3.6}$$

where k_{normal} and $x_{penentration}$ are stiffness in normal direction and displacement of penentration.



(a)

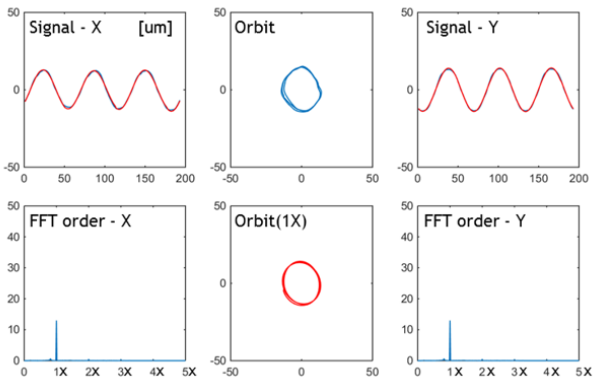


(b)

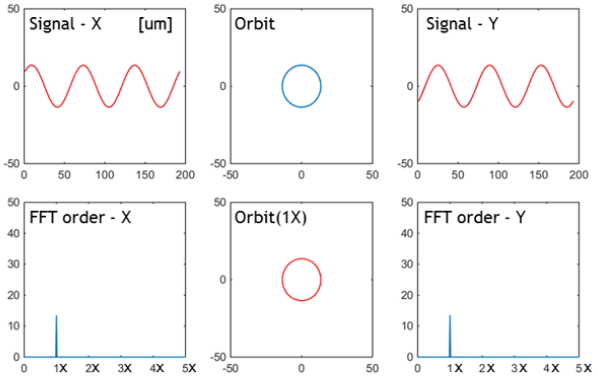
Figure 3-5 FEA model for rotors (a) schematics of target system (b) ANSYS model

Dynamic responses of normal and rubbing conditions at 3,600 RPM were obtained using transient analysis for two seconds with 0.001s interval. Especially two levels of unbalance were employed to validate the linearity between the magnitude of unbalance and the amplitude of vibration. The results of normal condition are shown in Figure 3-6 and Figure 3-7. The figure shows time-base signal, orbit and FFT. The blue line and red line in the figure means total signal and 1x component signal. Behaviors of rotor system in test-bed and FEA model show

similar results, and the linearity between unbalance and vibration magnitude are well represented. Figure 3-8 shows the results of rubbing condition in test-bed and FEA model. The case of impact rubbing was interpreted by contacting screw in test-bed and penalty stiffness in FEA model. The super-harmonic component in frequency analysis and truncated signal in time-domain which are main phenomena of impact rubbing were well represented in the results of test-bed and model.

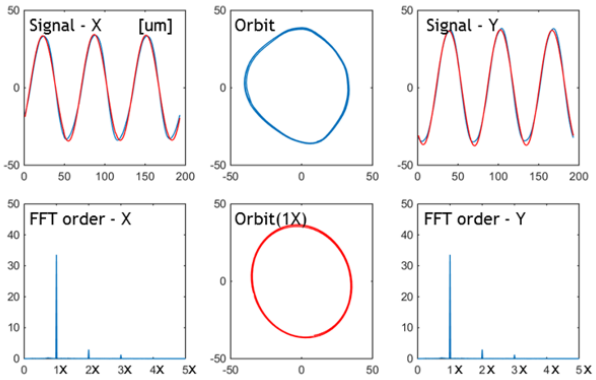


(a)

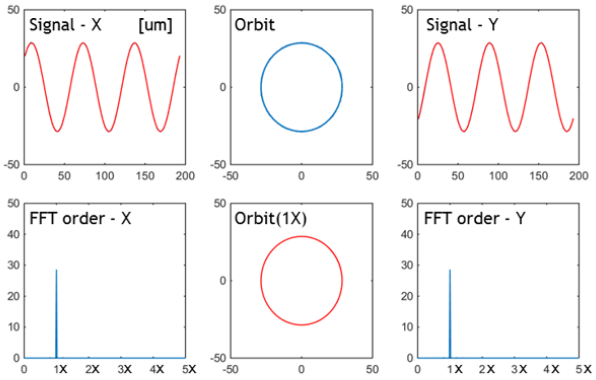


(b)

Figure 3-6 Normal states for 5 μm in RMS from (a) test-bed (b) FEA model



(a)



(b)

Figure 3-7 Normal states for 10 μm in RMS from (a) test-bed (b) FEA model

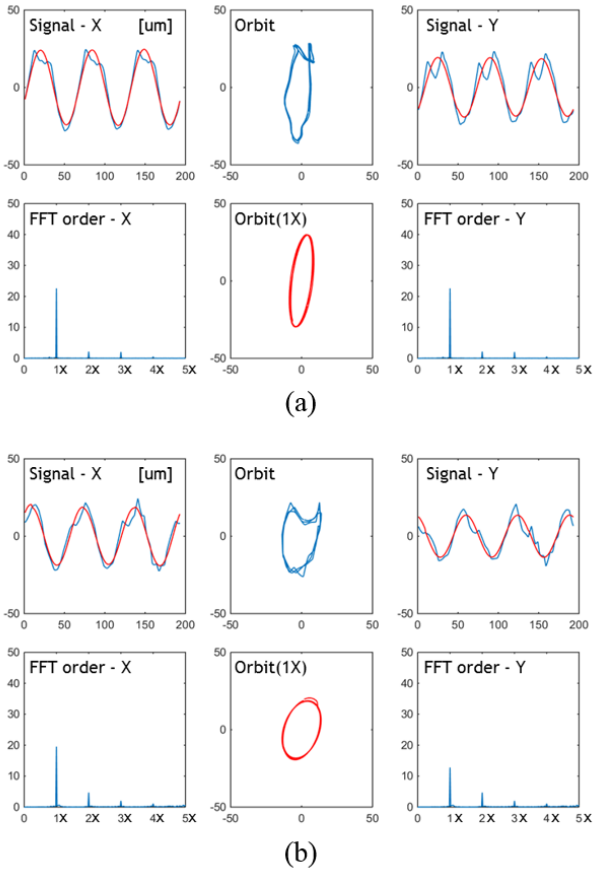


Figure 3-8 Impact rubbing conditions from (a) test-bed (b) FEA model

3. 2 Diagnosis Module for a Journal Bearing Rotor System

Supervised machine learning has been widely used as a method for system diagnosis in journal bearing systems. In this section, a common diagnosis procedure is described where a preprocessing technique, known as angular resampling,

minimizes the randomness in gap signals. Subsequently, the candidates for health features are described in both time- and frequency-domains.

3.2.1 Diagnostics Procedures

A diagnosis module must be carefully designed to assure reliable operation of the system by anticipating potential failures of the system. A supervised learning method is suitable for the study described here because training data sets can be acquired from rotor test-beds or a power plant turbine rotors. The trained data sets gathered under normal and various anomaly conditions facilitate the development of the supervised learning process. The supervised algorithm can then be used to diagnose the system in operation. The overall procedure is presented in Figure 3-9.

The first step is to acquire training data from a rotor test-bed or from a power plant turbine rotors under normal and anomaly conditions. As discussed, a gap signal is used for diagnosing a journal bearing rotor system. Features representing anomaly states must be carefully extracted from the signals obtained through time, frequency,

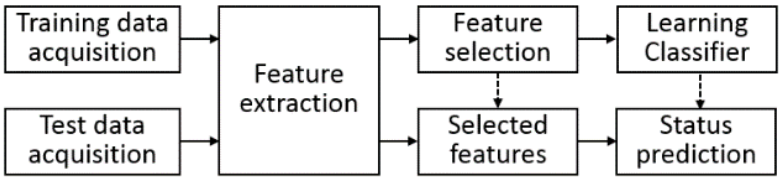


Figure 3-9 Data-driven diagnosis procedure

and/or time-frequency analysis. A feature selection algorithm can be used to select an optimal feature set for diagnosis efficiency [32]. The selected feature set can provide a logical base for accurate health classification of the journal bearing rotor system. Finally, the test data acquired from rotor systems in operation can be processed and classified into normal or one of the defined anomaly states using a trained classifier [114, 115].

3.2.2 Time- and Frequency-domain Features

Time- and frequency-domain features were more suitable than time-frequency domain ones because turbine rotors operate at steady-state operational conditions (i.e., 3,600 rpm). Table 3-2 lists the time-domain features, such as maximum, RMS, kurtosis, etc. [30, 116]. The first three features imply kinetic energy of the rotor. The next two features describe the statistical characteristics of the gap signals. The last three features indicate the shape of the sinusoidal wave. In addition, power spectrum data can represent important characteristics of vibration signals, which can be categorized as frequency domain features. The description of frequency features is shown in Table 3-3 [29, 117] where f and s(f) denote the frequency and the power spectrum function, respectively. As defined in Table 3-3, frequency center (FC) and root mean square frequency (RMSF) are related to the fundamental frequency of the system. The root variance frequency (RVF) describes how well the power spectrum is grouped. Other features represent the ratios of harmonic frequencies to the

fundamental frequency. These frequency features are commonly used in characterizing anomalies.

Table 3-2 Time-domain features

Notation	Features	Description	Physical interpretation
T1	Maximum	$\mathrm{Max}(X_i)$	Kinetic energy related
T2	Absolute Mean	$\mathrm{Mean}(\lvert X_i \rvert)$	
T3	RMS	$\sqrt{\frac{\sum {X_i}^2}{N}}$	
T4	Skewness	$\frac{\sum (X_i - \bar{X})^3}{(N-1)s^3}$	Data statistics related
T5	Kurtosis	$\frac{\sum (X_i - \bar{X})^4}{(N-1)s^4}$	
T6	Crest Factor	$\frac{X_{peak}}{X_{rms}}$	Sinusoidal wave shape related
T7	Shape Factor	$\frac{X_{rms}}{\mathrm{Mean}(\lvert X_i \rvert)}$	
T8	Impulse Factor	$\frac{\mathrm{Max}(X_i)}{\mathrm{Mean}(\lvert X_i \rvert)}$	

Table 3-3 Frequency-domain features

Notation	Features	Description	Physical interpretation
F1	FC	$\frac{\int f \times s(f)df}{\int s(f)df}$	Position change of main frequencies
F2	RMSF	$\left[\frac{\int f^2 \times s(f)df}{\int s(f)df}\right]^{1/2}$	
F3	RVF	$\left[\frac{\int (f-FC)^2 \times s(f)df}{\int s(f)df}\right]^{1/2}$	Convergence of spectrum power
F4	0.5X / 1X	$\sqrt{\frac{s(f_{0.5X})}{s(f_{1X})}}$	Magnitude of the certain frequency range
F5	2X / 1X	$\sqrt{\frac{s(f_{2X})}{s(f_{1X})}}$	
F6	(3x~5x)/1x	$\frac{\sum_{n=3}^{5}\sqrt{s(f_{nX})}}{\sqrt{s(f_{1X})}}$	
F7	(3x,5x,7x,9x)/1x	$\frac{\sum_{n=1}^{4}\sqrt{s(f_{(2n+1)X})}}{\sqrt{s(f_{1X})}}$	
F8	(2x~10x)/1x	$\frac{\sum_{n=2}^{10}\sqrt{s(f_{nX})}}{\sqrt{s(f_{1X})}}$	
F9	(0-0.39x)/1x	$\frac{\int_{0}^{0.39X}\sqrt{s(f)}\,df}{\sqrt{s(f_{1X})}}$	
F10	(0.4x-0.49x)/1x	$\frac{\int_{0.4X}^{0.49X}\sqrt{s(f)}\,df}{\sqrt{s(f_{1X})}}$	
F11	(0.51x-0.99x)/1x	$\frac{\int_{0.51X}^{0.99X}\sqrt{s(f)}\,df}{\sqrt{s(f_{1X})}}$	

3.2.3 Feture Selection Problem

The goal of the feature selection is to determine an optimal set of features that best represents the system health states. The optimization problem is formulated to attain the optimal feature set by maximizing the class separation function (PoS) and minimizing the correlation coefficient between features to eliminate redundant features, if any.[118, 119] For a given size (m) of the feature set, the optimization problem can be formulated as:

$$\text{Maximize} \quad \Pi_i = \frac{1}{m}\sum_{j=1}^{m} PoS_j - \frac{1}{{}_mC_2}\sum_{j\neq l}^{m}\left|\rho_{j,l}\right| \tag{3.7}$$

where PoS_j is the PoS of j^{th} feature and $\rho_{j,l}$ is the correlation coefficient between the j^{th} and l^{th} features. The design variables to be optimized are given as:

$$\boldsymbol{x}_d = [x_1, x_2, \dots, x_m] \tag{3.8}$$

where $\boldsymbol{x}_d$ is a candidate feature set composed of m number of features.

Figure 3-10 displays the overall feature selection procedure using the GA. In this procedure, the size of the feature set begins with $m = 2$ and gradually increases up to the maximum number of the candidate features. The optimum feature set for each set size can be determined by maximizing the cost function (Π_i) using the GA. In order to find the optimal solution, a population, N_p number of candidate feature sets is randomly generated at first, and the cost function for each feature set in the

population is calculated. Then, two criteria are used to decide the termination of the optimization. One is the lower limit criterion (β) of the difference between the maximum and the average value of Π_i of the population. If the criterion is not reached, the population is modified using the GA to meet the criterion. A sequence of modification can be regarded as one generation. The other criterion, k_{max}, is the upper limit of the generation numbers. The algorithm can stop at the k_{max} generation step even if the first criterion is not satisfied. A set from the last generated population, which has the maximum value of Π_i, is selected as the optimal set. In this study, N_p, β, and k_{max} are set to $200\times m$, 0.02, and 80, respectively, as presented in Ref. 48. Note that, in updating the feature sets using the GA, the crossover operator BLX-α[120] in equation (11) is used to select a better feature subset. The GA feature selection can be expressed as:

$$g_c^n(k+1) = round(\alpha \cdot g_1^n(k) + (1-\alpha) g_2^n(k)) \tag{3.9}$$

where k is the generation number, $\alpha \in [0,1]$ with uniform distribution, g_c^n means the n^{th} feature of the updated feature set c, and g_1^n and g_2^n represent the n^{th} feature of the parent feature sets. The optimal feature sets for a given size (m) of the feature set are used to demonstrate the effectiveness of the optimal datum units suggested by the PoS measure, as described in Section 4.3.3

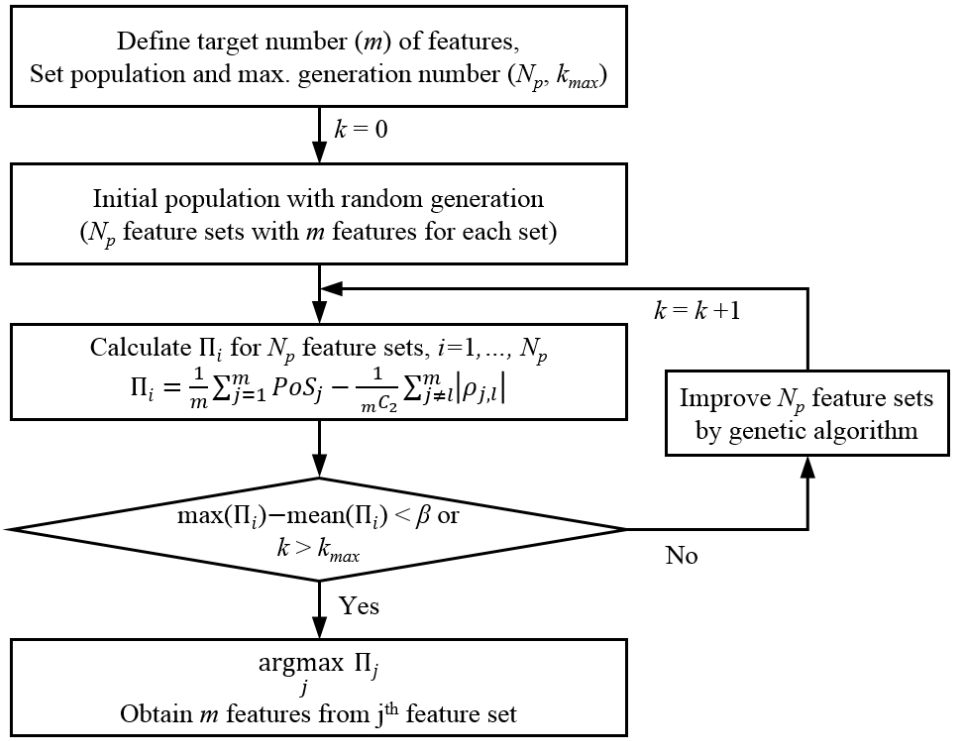


Figure 3-10 Feature selection procedure using the GA

3.2.4 Support Vector Machine (SVM) Classifier

In a supervised learning, each state of training data is labeled as an individual class. Classification algorithm identifies the class of the testing data by a trained classifier. Training features extracted from training data are used to define statistical characteristics for learning classifiers. By using the trained statistical rules, the testing features extracted from testing data are classified into one of the trained classes..

Health classification is accomplished through the learning of the classifier using the optimal features. In this study, the SVM classifier was used since it generally outperforms other classifiers. SVM is a supervised learning method that constructs an optimal hyper-plane by solving the structural risk minimizing function. In a two-class case, the training data set can be given as follows:

$$\{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \dots, (\boldsymbol{x}_{\mathrm{m}}, y_{\mathrm{m}})\} \tag{3.10}$$

where $\boldsymbol{x}_i \in R^n$ and $y_i \in \{1, -1\}$. $\boldsymbol{x}_i$ is the i^{th} n-dimensional training data, and y_i is the class index of the i^{th} training data. The general form of a hyper-plane for linear discrimination can be described as:

$$\boldsymbol{w} \cdot \boldsymbol{x} - b = 0 \tag{3.11}$$

where $\boldsymbol{w}$ is the normal vector of the hyper-plane, and b is the bias value. The distance between the two support vectors from each class can be expressed as:

$$\frac{2}{\|\boldsymbol{w}\|} \tag{3.12}$$

Thus, the optimal hyper-plane is defined by solving the following optimization problem as:

$$\begin{aligned} &\text{minimize} && \frac{\|w\|^2}{2} + C\sum_{i=1}^{m}\xi_i \\ &\text{subject to} && y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i - b) \geq 1 - \xi_i \end{aligned} \tag{3.13}$$

where ξ_i is a slack variable and C is a penalty coefficient. A slack variable is a non-negative value that allows a soft margin. A penalty coefficient controls the complexity of the hyper-plane and the classification error rate. In addition, if kernels are applied, a linear classifier turns into a non-linear one. The general description of a kernel, K, can be written as:

$$K(x_i \cdot x_j) = \Phi(x_i) \cdot \Phi(x_j) \tag{3.14}$$

where Φ is the transform function. As a result, the final decision function is expressed as:

$$\operatorname{sgn}\left(\sum_{i=1}^{m} y_i \alpha_i \cdot K(x_i \cdot x_j) + b\right) \tag{3.15}$$

In order to solve a multi-class problem, several binary SVM classifiers are required. There are two primary approaches, One-Against-One (OAO) and One-Against-All (OAA). In this research, the OAO method that Hsu and Lin suggested was used because it was found to be a superior approach for practical use.[89] This research implemented commercial SVM code known as "LIBSVM[121]" with a linear kernel function.

Chapter 4. Methodology for Datum Unit Optimization

While ongoing research has improved diagnostics efforts, one of the remaining issues in feature extraction is the determination of the optimal amount of data for feature generation. Prior work has shown that the sampling rate of data acquisition and the amount of data available for diagnosis have an effect on the performance of the diagnosis algorithm [47, 48]. This implies that an optimum datum unit must be decided with much care for robust fault diagnosis of rotor systems, even after the signal is resampled. Thus, research is needed to determine the optimum datum unit—the focus of this study.

Feature selection is a methodology to find an optimal set of features that best represents the states and to exclude irrelevant or redundant features. It generally contains an optimization problem in order to select the best features using a separability measure as a cost function. Kullback-Leibler divergence (KLD) has been proposed as a separability measure [122]. KLD quantifies the dissimilarity between two probability density functions (PDFs) for different features. The Fisher discriminant ratio (FDR), another separability measure, represents the distance between two PDFs using scatter matrices of mean and variance [123]. KLD has the advantage of estimating separability close to a Bayes error when the PDFs are known exactly. FDR is a straightforward separability measure because it only uses the mean and variance. However, these two measures have limitations in that they have no

upper limit for a fully separable case. A measure newly proposed in this paper, probability of separation (PoS), can be used to overcome this disadvantage. PoS quantifies the degree of class separation in a way that is analogous to the idea of load-strength interference [124]. The research described in this paper found that PoS outperforms KLD and FDR as a separability measure in a data-driven approach.

4.1 Pre-processing for Gap Sensor Signals

Although the test-bed is set to operate at a target speed, for example 3,600 rpm, the speed actually varies randomly. The variation is no more than 20 rpm during normal operation of the test-bed, but higher rpm alterations can be observed under a sudden impact or during an abrupt change. In general, turbine rotor systems may have greater variation of rpm because there are numerous uncertain factors of operating conditions such as an output power, steam pressure, oil temperature of the journal, etc. Low-pass filter which has cutoff frequencies in 400 Hz was implemented to reduce the effect of high frequency noise. Then the angular resampling process is applied for extracting accurate features amidst the substantial noise in the signals. This resampling process regenerates the acquired data based on the rotation angles and facilitates robust diagnosis of a journal bearing rotor system [14-16]. The idea of angular resampling is outlined in Figure 4-1. This process enables the vibration data to have an equal angle difference between two adjacent data points in a rotation. As a result, regardless of the rpm variation, the resampled

signal can have an equal number of points per cycle, which increases reliability in extracting features. Figure 4-2 represents the comparison of dispersion of the raw signal and the angular resampled signal from the RK4 test-bed. The variance of the resampled signal is substantially reduced, which strengthens diagnostic performance. In this research, every single cycle between the neighboring keyphasor signals was resampled to have 128 points. Features were then extracted for analysis.

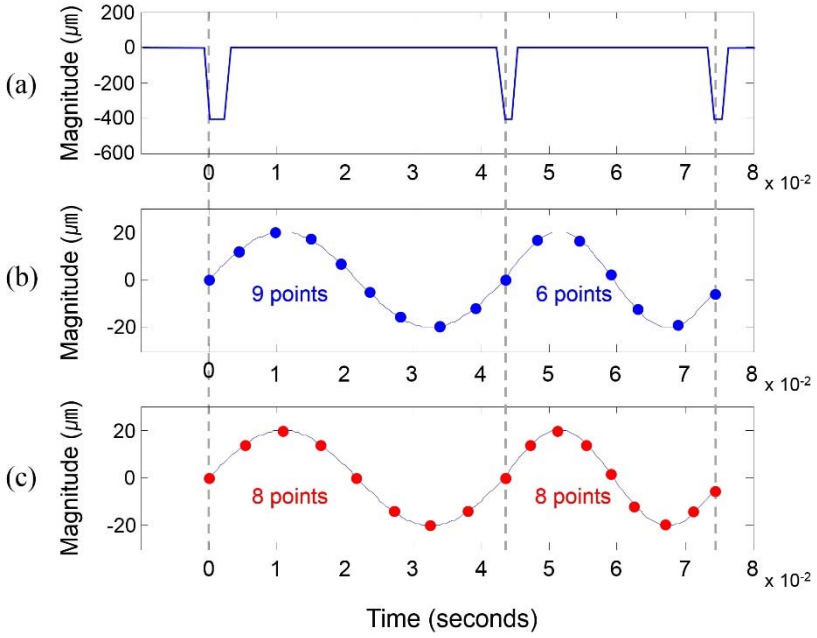
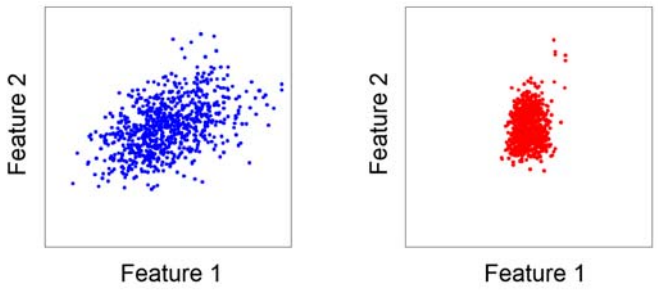


Figure 4-1 Resampling outline (a) keyphasor signal (b) raw signal (c) resampled signal



(a) (b)

Figure 4-2 Distributions of (a) raw data features and (b) resampled data features

4.2 Definition of Feasible Datum Units

The resampling process described in Section 4.1 can split the data by one complete cycle, which corresponds to a 360-degree of rotation angle. This technique enables features to be extracted based on the exact number of cycles. However, there still remains one last problem, specifically: how many cycles are optimal for a particular feature. The motivation for this study came from the idea that even a slight change when acquiring data or when processing data could result in different statistical characteristics. This section concentrates on defining the feasible datum units for feature extraction

The eight time-domain features used to obtain the statistical characteristics from vibration data are described in Table 3-2. The goal of this research was set to define the optimal datum units. To assess the effect of the number of cycles on feature characteristics, the eight time-domain features were extracted based on 1, 10, 20, 30,

and 60 cycles. The same amount of data using the datum units of 1, 10, 20, 30, and 60 cycles are employed for comparison between various datum units. The maximum of feasible data size must be decided by considering the performance of data acquisition devices and the computational ability of the software in processing the signals. Sixty-cycles, which is equivalent to the length of a one-second signal, was thus set as the maximum size of a datum unit. The datum units of sixty-cycles and divisors of sixty-cycles were compared in terms of their diagnosis performance. Some submultiples were excluded to avoid excessive computational loads.

The eleven frequency-domain features used to get information about vibration data are listed in Table 3-3. Initially, the feasible datum units defined for time-domain analysis were taken into account for extracting the frequency-domain features. Some frequency-domain features are sensitive to the number of cycles for feature extraction. That is because the resolution of the frequency response of a signal relies on the number of cycles. For example, the k-cycle datum unit has a resolution of $1/k$ times the fundamental frequency. It is well known that a few-cycle datum unit cannot represent the sub-harmonics of the fundamental frequency [51]. Thus, this research excluded the one-cycle datum for frequency-domain analysis.

4.3 Class separability Metrics

This section introduces three evaluation metrics to measure the capability of class separation. The metrics are used to assess the effect of the datum units on the capability of class separation and to suggest the optimal datum units for time- and frequency-domain features. The three evaluation metrics include: Kullback-Leibler divergence (KLD), Fisher discriminant ratio (FDR) and a new measure, namely Probability of Separation (PoS), proposed in this paper. Unlike the others, PoS quantifies the degree of class separation using a normalized scale, 0 to 1. This section discusses the merits and limitations of each of these three class separability metrics. All three metrics are used to determine the optimal datum units.

4.3.1 Kullback-Leibler Divergence (KLD)

The Kullback-Leibler divergence, also called *relative entropy*, is a widely used distance measure in a data-driven approach. Here, KLD was used to quantify dissimilarity between two probability density functions (PDFs) of a specific feature where each function denotes a condition. Equation (1) represents the definition of KLD for two continuous PDFs, $f(x)$ and $g(x)$ [122]. KLD is basically an asymmetric measure according to the order of the PDF, as stated in equation (1). Therefore, a symmetric KLD form in equation (2) was implemented as a means to quantify the degree of class separation [122]. Also, KLD can be represented in equation (3) for

the probability mass function P and Q in a discrete version. The KLD value is "0" when the two distributions are perfectly overlapped. In contrast, a high KLD value guarantees a good capability for class separation. KLD is closely related to the Bayes error, but it is often difficult to estimate for non-Gaussian distributed data [125]. In addition, KLD can be inadequate for comparison of the separability between features because different features may have different KLD values for the same degree of class separation [126, 127].

$$D_{KL}(f||g) = \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{g(x)} dx \tag{4.1}$$

$$\begin{aligned} D_{KL}(f,g) &= D_{KL}(f||g) + D_{KL}(g||f) \\ &= \int_{-\infty}^{\infty} \{f(x) - g(x)\} \log \frac{f(x)}{g(x)} dx \end{aligned} \tag{4.2}$$

$$D_{KL}(P,Q) = \sum_{i} \{P(i) - Q(i)\} \log \frac{P(i)}{Q(i)} \tag{4.3}$$

4.3.2 Fisher Discriminant Ratio (FDR)

The definition of FDR in a two-class problem is shown in equation (4), which uses only the mean and the variance.[123] μ_i and σ_i represent the mean and standard deviation of i^{th} class. The numerator indicates the difference between the two-class means while the denominator normalizes the variances of the two classes. Given a

relatively large separation of two-class data, the FDR results in a great value. FDR is easy to use since it only uses the mean and variance for each class. On the other hand, it may provide incorrect information when non-normal distributions are engaged.

$$FDR = \frac{(\mu_i - \mu_j)^2}{\sigma_i^2 + \sigma_j^2} \tag{4.4}$$

4.3.3 Probability of Separation (PoS)

PoS is also a two-class separability measure based on the idea of load-strength interference. Failure occurs when an applied load exceeds the strength (e.g., fracture toughness) of a mechanical part. Taking into account uncertain load and strength variables, the probability of failure can be described as the chance that the load variable exceeds the strength.[128] The equation of probability of failure is shown as:

$$P_f(s) = \int_{-\infty}^{\infty} f_L(s) F_S(s) dx \tag{4.5}$$

Analogous to the load-strength interference, the probability of a non-separable region can be defined as:

$$P_{NS} = \int_{-\infty}^{\infty} f_{c1}(x) F_{c2}(x) dx \quad \text{for} \quad \tilde{x}_{c1} \leq \tilde{x}_{c2} \tag{4.6}$$

where $f_{c1}(s)$ and $F_{c2}(s)$ represent the probability density function (PDF) of class 1 and cumulative distribution function (CDF) of class 2, respectively, while $\tilde{x}_{c1}$ and $\tilde{x}_{c2}$ correspond to the medians of classes 1 and 2. Noting that the probability of a non-separable region ranges between 0 and 0.5, PoS is defined as the normalized the separability value between 0 and 1. The equation of PoS is expressed as:

$$PoS = (e^{(1-2\times P_{NS})} - 1)/(e - 1) \tag{4.7}$$

PoS gives "0" if the feature data of two different classes overlaps perfectly and "1" if not overlapped at all. The bounded and normalized property makes PoS more powerful than either of the other two separability measures for partially separated data of two classes, which is the most common scenario in the classification problem. Some comparative study will be discussed in the following section.

4.3.4 Discussion on the Measures of Class Separation

A comparative study among three class separation measures was performed for two-class separation, as exemplified in Figure 4-3. This study entailed four separation cases: no separation, a minor separation, a major separation, and full separation. The gap signals measured from the RK4 rotor test-bed were used for generating the feature data. Due to the randomness in the feature data, the histogram

for each feature was employed for evaluating the degree of class separation using the separability metrics in a two-class problem. When two-class problems are expanded to a multi-class problem, an average of the separation values for all possible combinations of two-class problems represents the separation value in the multi-class problem, as shown in equation (8). S_{M_c} and $S_{i,j}$ represent the separability value for the M-class case and for the two class case in i^{th} and j^{th}, respectively. $S_{i,j}$ can be obtained by using one of the three measures—KLD, FDR, or PoS. Four health classes entail six (= ${}_4C_2$) possible combinations.

$$S_{M\_c} = \frac{1}{{}_M C_2} \sum_{i}^{M} \sum_{j \neq i}^{M} S_{i,j} \tag{4.8}$$

Table 4-1 summarizes the results of class separation. Some observations on the performance of each of the class separation measures are summarized as follows:

(i) In the case of no separation, which is shown in Figure 4-3 (a), all three measures have the class separation value of "0."

(ii) In the case of full separation, which is shown in Figure 4-3(d), different values are assigned by each of the three measures. KLD and FDR describe the full separation inconsistently, depending on the distribution shape. Unlike the earlier methods, PoS always gives the class separation value of "1" for the perfect separation.

(iii) In the case of partial separation, PoS expresses the degree of class separation in a normalized scale [0-1] proportional to how much the two distributions are distinct. PoS values of 0.752 and 0.989 state the degrees of class separation for the cases shown in Figure 4-3(b) and (c), respectively. KLD and FDR quantify the class separation in a non-normalized scale. KLD is more sensitive in between minor (Figure 4-3(b)) and major separations (Figure 4-3(c)), whereas FDR is more sensitive in between major (Figure 4-3(c)) and full separation (Figure 4-3(d)).

Figure 4-4 and Table 4-2 show clearly that KLD and FDR can give counterintuitive results because they depend on a bin size and distribution normality. Figs. 8(a) and (b) have the same degree of class separation with different bin sizes. KLD gives different separation values, unlike the other measures. This implies that KLD is bin-size dependent. For this very reason, KLD may result in another counterintuitive situation, as shown in Figure 4-4(b) and (c). FDR may also mislead the class separation process because the FDR measure cannot deal with non-normally distributed histograms properly, as shown in Figure 4-4(c) and (d). In summary, KLD and FDR must be used with great care; otherwise, they may eventually produce incorrect diagnosis results.[126, 127]

As summarized in Table 4-3, PoS has three favorable properties: (a) normalization, (b) boundedness, and (c) independence from histogram. These attributes make the proposed measure, PoS, far more favorable as a class separation measure. PoS expresses the degree of class separation in a normalized scale, which

allows an intuitive interpretation of the class separation. Moreover, the boundedness property in PoS always assigns the minimum (=0) and maximum (=1) values of PoS to no and fully separated cases, respectively. When building a histogram, it may not be normally distributed, or its bin size can be arbitrary chosen. In either case KLD and FDR can be affected, whereas PoS is not.

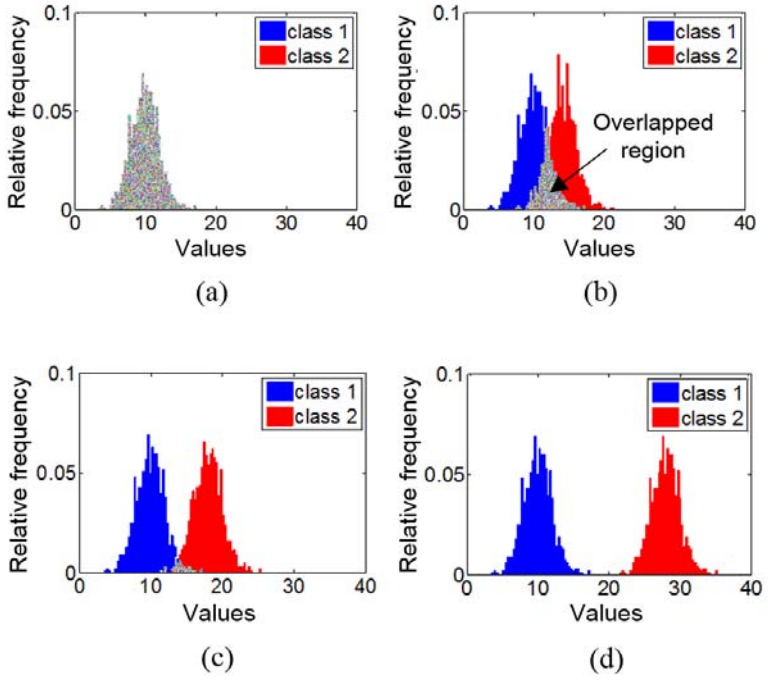


Figure 4-3 Four scenarios of class separation

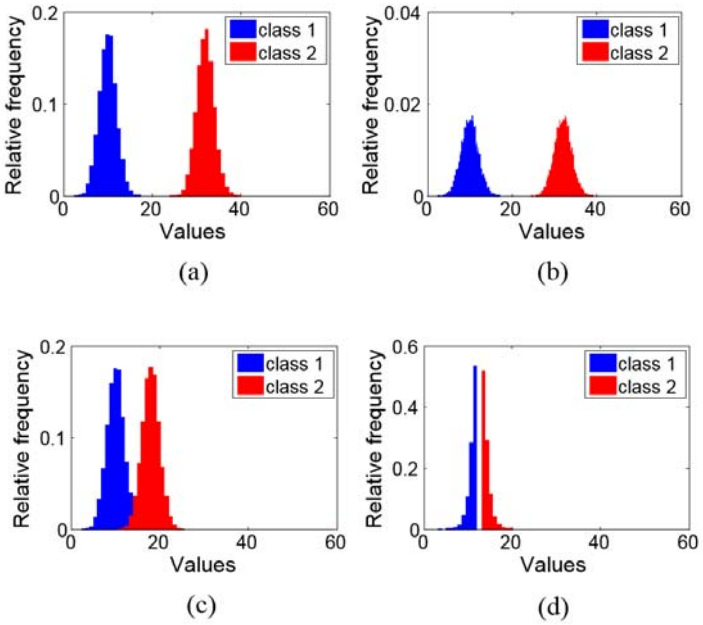


Figure 4-4 Dependency on a bin size and distribution normality

Table 4-1 Class separation results in Figure 4-3

	(a)	(b)	(c)	(d)
KLD	0	6.589	20.278	23.692
FDR	0	2.004	8.017	40.584
PoS	0	0.752	0.989	1.000

Table 4-2 Class separation results in Figure 4-4

	(a)	(b)	(c)	(d)
KLD	26.801	19.701	20.506	21.117
FDR	61.545	61.545	8.138	5.187
PoS	1.000	1.000	0.992	1.000

Table 4-3 Summary of the advantages and limitations for the three separability measures

Criteria	Advantages	Limitations
KLD	Close to Bayes error with exact PDFs	Not bounded Bin size dependent
FDR	Simple to implement and use Bin size independent	Not bounded Not applicable to non-normally distributed histogram
PoS	Normalized Bounded Histogram independent	

4.4 Diagnosis Results via Various Datum Units

As mentioned above, the focus of this study is on defining the optimal datum units for vibration data that best strengthen the diagnostic performance of a journal bearing system. The feasible datum units were comparatively studied using several qualitative techniques, as described in Section 4.4.1. Section 4.4.2 suggests the optimum datum units through quantitative evaluation of the feasible datum units for anomaly diagnosis of a journal bearing system. Section 4.4.3 describes how the optimal datum units were verified by assessing the prediction accuracy of health classification.

4.4.1 Qualitative Study of Anomaly Diagnosis

A qualitative study for anomaly diagnosis is essential. Time-, frequency-, and time-frequency analyses are often employed for examining anomaly states using vibration signals, i.e., acceleration, gap. In this research, gap signals were acquired from the RK4 test-bed under four health states: normal, rubbing, misalignment, and oil whirl.

Empirical histograms of T6 and F2 acquired from RK4 experiments characterize the four classes, as shown in Figure 4-5 and Figure 4-6. The one-cycle datum unit is excluded for the frequency-domain analysis because the features (F4, F9, F10 and F11) representing sub-harmonics cannot be obtained with the one-cycle datum unit. The effect of datum units on class separability can be visually evaluated from the figures. Comparing the dispersion of the relative frequency in Figure 4-5, it is shown that an increased size of a datum unit leads to a greater variance of the T6 feature. In particular, the one-cycle datum unit of the oil whirl case is exceptionally different. In Figure 4-6, the oil whirl case exhibits a little variation of the F2 feature over the feasible datum units, while the others indicate little difference over the datum units. It is observed that the frequency resolution determined by the datum unit influences the F2 value of the oil whirl case.

Although Figure 4-5 and Figure 4-6 present a qualitative analysis which gives intuitive and easy understanding of the class separability, a quantitative study is

required for more precise determination of the optimal datum units. The following section covers the quantitative study for anomaly diagnosis, which provides a logical basis for selecting the optimal datum units.

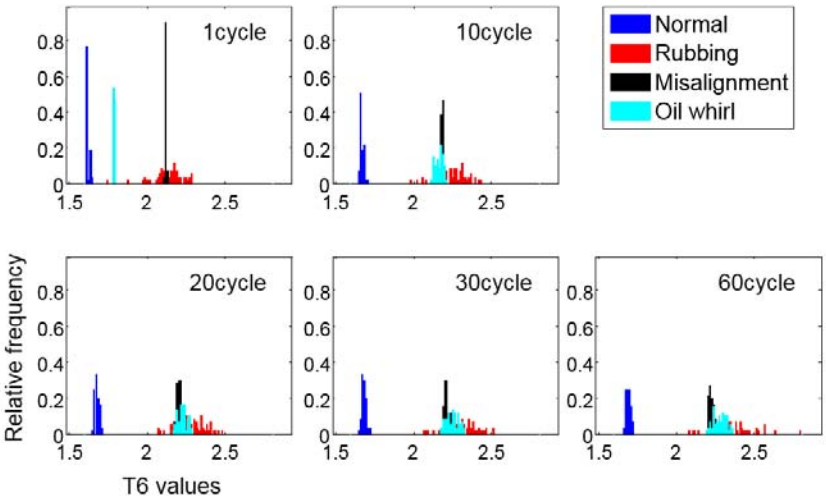


Figure 4-5 Distributions of the T6 feature over the feasible datum units

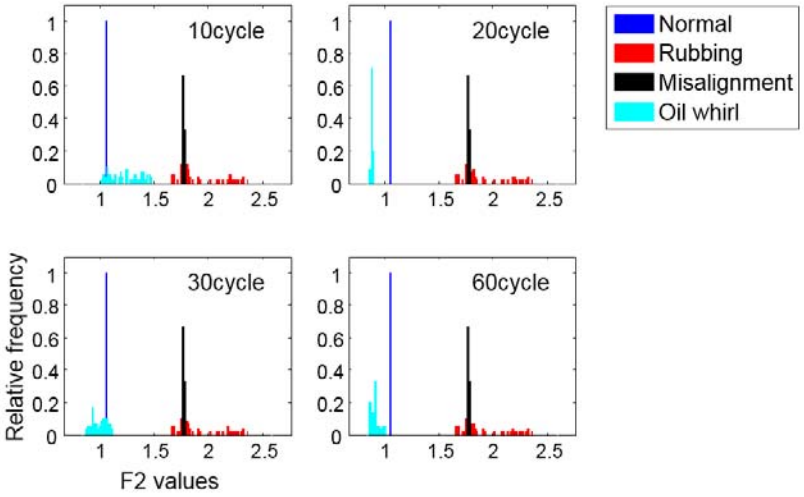


Figure 4-6 Distributions of the F2 feature over the feasible datum units

4.4.2 Quantitative Study of Anomaly Diagnosis

This section discusses the performance of class separation using three evaluation metrics, specifically KLD, FDR, and PoS. This study used the vibration signals from RK4 experiments to characterize the four health conditions. Each of the eight time-domain features and the eleven frequency-domain features were calculated with different datum unit sizes. A separability measure for multi-class cases was obtained from an average of separability measures for all possible combinations of two-class problems, as shown in equation (8). The separability results of time- and frequency-domain features using the three evaluation metrics are shown in the following sections.

Time-domain Features

KLD values for time-domain features are shown in Figure 4-7(a). A one-cycle datum unit was observed to be superior in most time-domain features compared to the other datum units. For the time-domain features that are obtained through the integration of data—T2 (mean), T3 (RMS), and T7 (shape factor), the separability values appear to be steady over the feasible datum units. This implies that the datum size is not important for the data-integrated time-domain features. This trend is observed in both FDR and PoS methods. On the other hand, the KLD values of T6 (crest factor) and T8 (impulse factor) were reduced as the datum unit size increased.

The reduction in the KLD values was due to the increased variance of features, which can be observed in Figure 4-5.

The results of the FDR values are shown in Figure 4-7(b). Four time-domain features (T1, T4, T6, T8) of the one-cycle datum unit acquired the highest FDR values. The high FDR values result from the relatively small variances compared to those of the other cycle datum units, as shown in Figure 4-5. Just like KLD, three

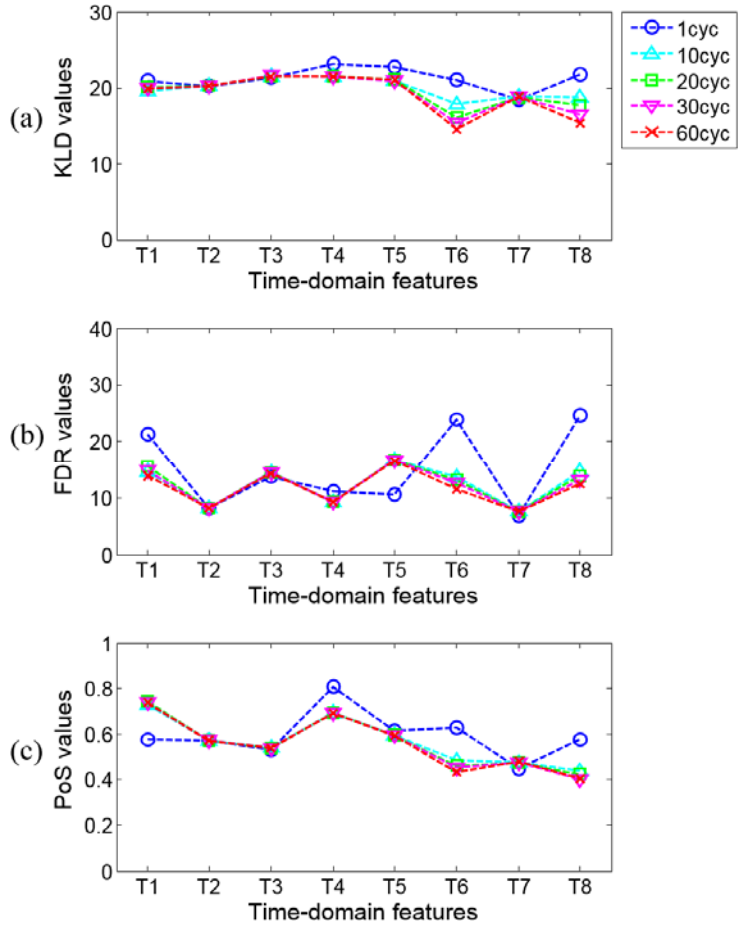


Figure 4-7 (a) KLD (b) FDR (c) PoS separability values of time-domain features

other time-domain features (T2, T3, T7) showed the same or similar FDR values over all datum unit sizes. However, the T5 (Kurtosis) feature gave the opposite result. In this case, the one-cycle datum unit obtained the lowest FDR value because, in the one-cycle datum unit, the T5 feature distribution of oil whirl was shifted towards that of normal. The FDR value of the one-cycle datum unit was decreased due to the reduction in the mean difference between normal and oil whirl, although they are fully separated. As discussed in Section 4.3.4 and Table 4-3, KLD and PoS measures are less sensitive than FDR in fully separated cases.

Overall, the one-cycle datum unit earned the highest PoS in the time-domain features, as shown in Figure 4-7(c). However, the T1 (maximum) feature gave a conflicting result. For a sub-harmonic dominant signal, such as oil whirl, the maximum values for each single rotation cannot represent the global maximum of the period because the period is longer than one revolution. Therefore, T1 values for oil whirl of the one-cycle datum unit are reduced as compared to those found for the other datum units. This reduction caused the lowest PoS value for T1. However, KLD and FDR values of T1 gave inadequate results from a combination that had exceptionally high values because of non-boundedness.

In summary, the one-cycle datum unit offers a superior capability of class separation in most of the time-domain features. The selection of the datum sizes are not important for the time-domain features which are obtained through the integration of data—T2 (mean), T3 (RMS), and T7 (shape factor). Among the three

evaluation metrics, PoS is the most favorable because of its properties of normalization and boundedness.

Frequency-domain Features

The frequency-domain features are extracted from the results of power spectral density (PSD). Figure 4-8 shows the separability values of the frequency features over the feasible datum units. For the analysis of the frequency-domain features over the datum units, the features can be divided into two groups. The first group (F1-F3) includes the global features, which encompass the entire frequency content, whereas the second group (F4-F11) are the local features that are interested in a particular frequency content.

Let us first consider the first feature group, which contains the integration of both frequency and magnitude of PSDs. F2 (RMSF) and F3 (RVF) features showed greater variation over the datum units, regardless of the separability measure used. The large variances are mainly due to different distributions of features in the oil whirl case according to the datum units used, as shown in Figure 4-6 for the F2 feature. For a sub-harmonics dominant signal, such as oil whirl, the magnitudes of distributions for each frequency are highly influenced by the frequency resolution determined by datum unit size. If the dominant frequency of oil whirl can be exactly represented in a certain resolution defined by the datum unit size, the F2 and F3 features would have low variation and high separation. For example, the oil whirl of

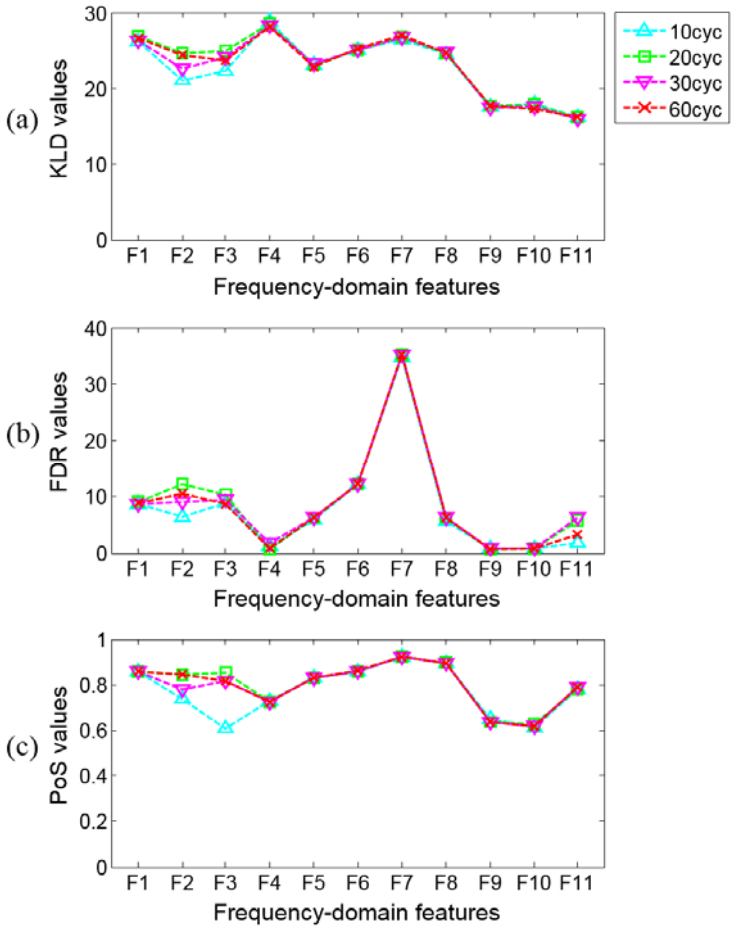


Figure 4-8 (a) KLD (b) FDR (c) PoS separability values of frequency-domain features

a 0.45x frequency component can be represented by the frequency resolutions of a twenty- or sixty-cycle datum unit. As a result, F2 and F3 features in twenty- and sixty-cycle datum units show high separability values, as shown in Figure 4-8. Similarly, another oil whirl case of a 0.43x frequency component can be represented by the frequency resolution of a thirty- or sixty-cycle datum unit. These frequency

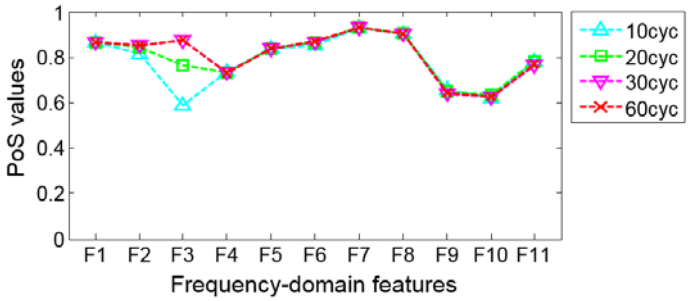


Figure 4-9 PoS separability values of frequency-domain features with inclusion of 0.43x dominant oil whirl signal

resolutions lead to high separability values in thirty- and sixty-cycle datum units, as shown in Figure 4-9. In addition, the F1 (FC) feature has the same trend as F2 and F3, but it has relatively small variances because the effect of frequency change in the equation for F1 feature is weaker than that of two other features. Further discussions related to frequency resolution are provided in Section 4.4.3.

The local feature group (F4-F11) offers nearly the same separability values over all datum units studied, regardless of the separability measure used. This result can be explained by the two sub-groups of the local features. The first sub-group (F9-F11) consists of the features extracted through the summation of magnitudes of PSDs for defined frequency ranges. The magnitude of a frequency component can be spanned to adjacent sidebands in certain datum units. However, the summation process reduces the variance of features over datum units by adding up the sidebands. The second sub-group (F4-F8) includes the features obtained from PSD magnitude value(s) at particular frequency contents. The particular frequencies used for the F4-

F8 features can be derived from the frequency resolutions of all datum units. Thus, each feature for the second sub-group has similar values across the various datum units because the magnitudes of PDFs are not affected by datum unit sizes.

FDR showed undesirable results due to its nature of non-normalization and unboundedness. This results in a greater variation in F11, substantially lower FDR values in the F4, F9, and F10 features, and conversely, a significantly higher FDR value in F7. This implies how important it is that the separability measure holds the properties of boundedness and normalization. Only PoS offers these features.

In the frequency-domain, a sixty-cycle datum unit, the highest frequency resolution, has been consistently supporting high separability, although most features are not affected by datum unit sizes.

4.4.3 Validation through Classification

This section describes the work performed to validate the effectiveness of the optimum datum units through the classification study. The SVM classifier is used for the classification of four health states implemented in the RK4 test-bed. Tests for each health condition were repeated twice. The training process used half of the data from each test (chosen randomly); the rest of the data was used in the prediction process. Classification studies were done with eight time- domain features and eleven frequency-domain features. Figure 4-10 shows the class prediction results

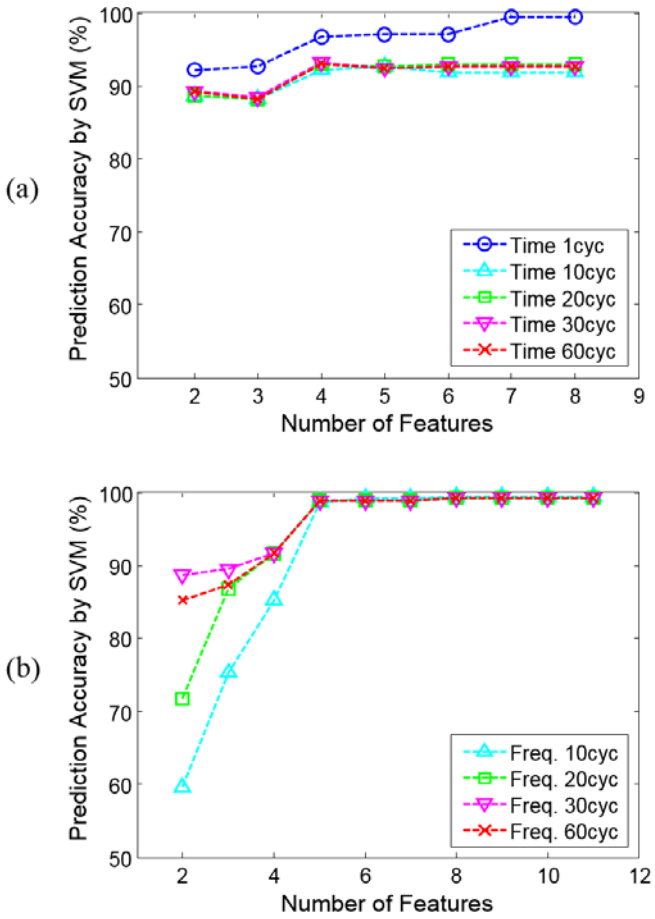


Figure 4-10 Classification results using (a) the time- and (b) the frequency-domain optimal features selected by the PoS-based GA

using the optimal features selected by the PoS-based GA. The numbers in the x-axis indicate the number of optimal features determined from the feature selection process. The classification was performed as the number of features increased.

The classification results of time-domain features are shown in Figure 4-10(a).

The increase in the optimal features of time-domain results in greater accuracy of class prediction. Moreover, the one-cycle datum unit was superior to the others from the viewpoint of classification accuracy. On the other hand, different datum unit sizes have no or little influence on the classification accuracy using the frequency-domain features, as shown in Figure 4-10(b). The accuracy becomes saturated with more than five optimal features regardless of the feature selection algorithm.

Another interesting issue is how the combination of the features in both the time- and frequency-domains affects classification accuracy. Figure 4-11 shows the classification results using the optimal features in both the time- and frequency-domains. A one-cycle datum unit was used for the time-domain, while all feasible datum units were used for the frequency-domain. A PoS-based GA was used for the

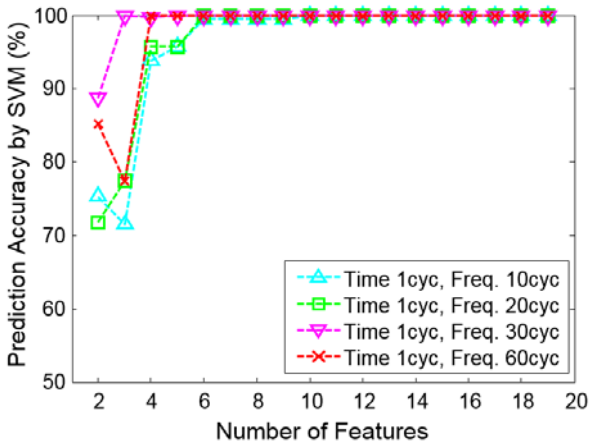


Figure 4-11 Classification results using both the time- and frequency-domain optimal features: A one-cycle datum unit for the time-domain and all feasible datum units for the frequency-domain

feature selection and the SVM classifier for the classification. 99% classification accuracy can be achieved using more than five optimal features in both the time- and frequency-domains. Among various datum units in the frequency-domain, thirty- and sixty-cycle datum units assure the maximum classification accuracy.

Some notable findings can be observed from the classification results:

- PoS-based feature selection can enhance the diagnostics performance.
- A one-cycle datum unit is optimal for time-domain classification.
- No single distinct datum unit was found for frequency-domain classification.

This section further discusses the third finding. It is well known that an oil whirl state produces sub-harmonic components, normally in the range of 0.42-0.48 times the fundamental frequency. The spectral leakage problem thus arises in the oil whirl state when the signal has a limited length.[129] This sub-harmonic component can be accurately represented by the twenty-, thirty- and/or sixty-cycle datum units, depending on the sub-harmonic components, where the twenty-, thirty- and sixty-cycle datum units have a resolution of 0.050x (=1/20), 0.033x (=1/30), 0.017x (=1/60) frequency, respectively. For example, twenty- and sixty-cycle datum sizes have no leakage for 0.45x frequency (= 9/20 and 27/60), whereas thirty- and sixty-cycle datum sizes have the minimum leakage for 0.43x frequency (≈ 13/30 and 26/60). Figure 4-12 and Figure 4-13 show the PSD and distributions of F1 over the various

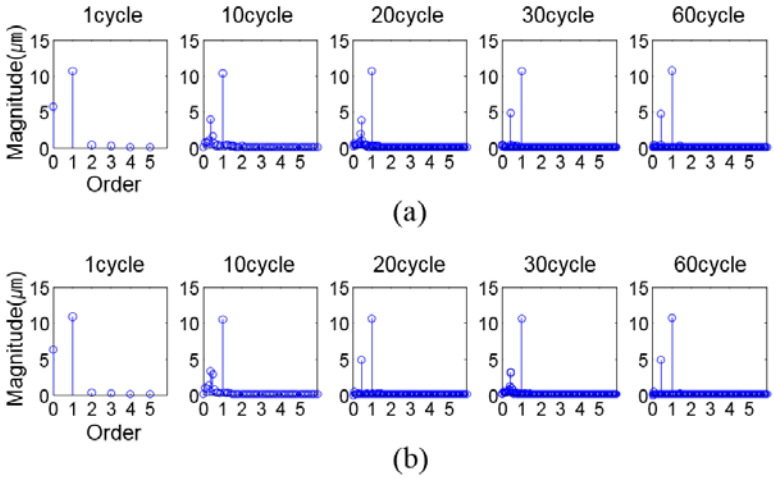


Figure 4-12 PSDs of oil whirl with (a) 0.43x and (b) 0.45x whirling frequency

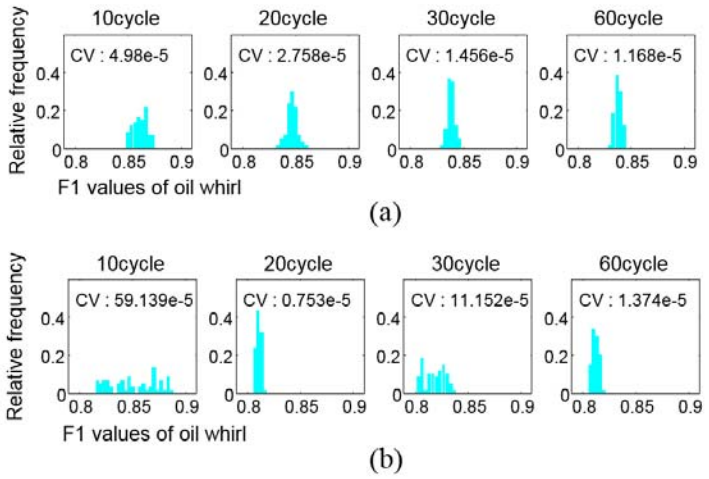


Figure 4-13 F1 feature distributions of oil whirl with (a) 0.43x and (b) 0.45x whirling frequency

datum units. Figure 4-12(a) and Figure 4-13(a) represent 0.43x frequency dominant oil whirl, while Figure 4-12(b) and Figure 4-13(b) exhibit 0.45x frequency

dominant oil whirl. The figures well represent that the small spectral leakage leads to the low variances of the F1 features. The variance of each distribution is quantified by the coefficient of variation (CV). It is uncertain which sub-harmonic component appears in the oil whirl state. As suggested in Section 4.4.2, it is thus desirable to select the sixty-cycle datum unit, which has the highest resolution among the feasible units. Of course, higher-cycle datum units can be used; however, they reduce the amount of feature data for a given signal. Unless signals are abundantly acquired and managed for anomaly diagnosis, the sixty-cycle datum unit is recommended for the frequency features.

Chapter 5. Omnidirectional Regeneration of Gap Sensor Signals

5.1 Omnidirectional Regeneration (ODR)

5.1.1 Definition

The acquired signals are obtained as a voltage that is inversely proportional to the gap between the sensor and the rotor. The voltage values are transformed into the exact distance by multiplying a scale factor, and eliminating the DC component of the voltage returns vibration signals. As the shaft rotates, as shown in Figure 5-1, the centerline changes from point 1 to 3 and the displacement between the rotor and the

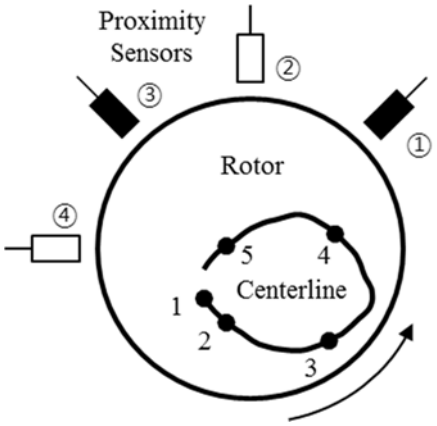


Figure 5-1 Sensor locations: ① and ③, and virtual sensor locations by ODR: ② and ④

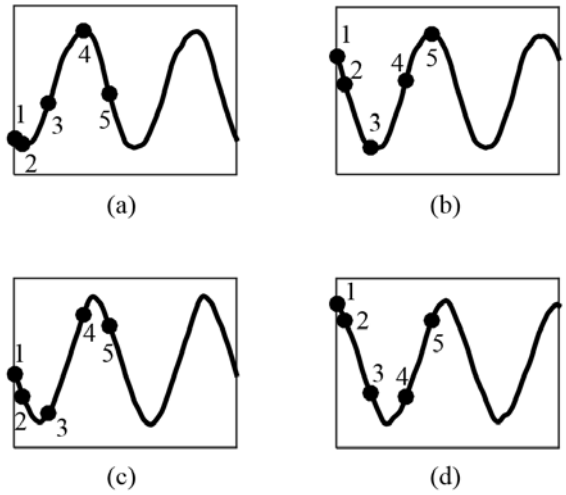


Figure 5-2 Vibration signals from sensor locations (a) ①, (b) ②, (c) ③, and (d) ④ in Figure 5-1

sensor ③ is increased. Accordingly, the vibration signals acquired by the sensor ③ in Figure 5-2 are decreased from point 1 to 3.

In addition, the vibration signals from ① and ③ sensors can make the orbit. Figure 5-3 represents the orbit made by the vibration signals. By using the vibration signals as the x- and y- coordinate of the Cartesian coordinate system, respectively, the orbit of the rotor can be obtained. The orbit drawn by the vibration signals exactly matches to the actual trajectory of the shaft centerline.

Omni-directional regeneration (ODR) is a method that produces vibration signals in an arbitrary direction. The vibration signals can be generated from the two

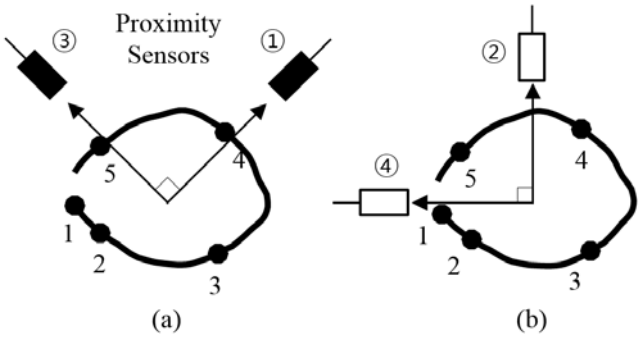


Figure 5-3 Orbits by signals from (a) ① and ③ sensors, and (b) ② and ④ sensors

vibration signals acquired by two orthogonal proximity sensors, as presented ① and ③ in Figure 5-1. The signals from ① and ③ are measured signals, while the signals from ② and ④ are generated by ODR. The corresponding signals are presented in Figure 5-2.

The omnidirectional regeneration (ODR) of vibration signals can be regarded as the signals obtained from an artificial sensor as presented in Figure 5-1 ② and ④. The ODR signals can be obtained by transforming the coordinates. The rotation about the origin of the Cartesian coordinate system generates signal which can be regarded as the measured signal at the rotated direction. In other words, the ODR is equivalent to placing an artificial sensor at a target direction. Thus, vibration signals in any direction that is of interest can be obtained if the real sensors are in orthogonal position.

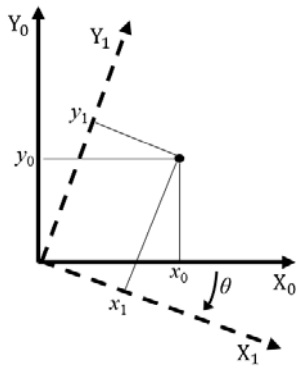


Figure 5-4 Coordinate transformation of a point in two-dimensional system.

The principle of ODR is to apply the coordinate rotation in the Cartesian system on two-dimensional plane. A data point *A* can be presented in a Cartesian coordinate as in Figure 5-4. The original coordinates, X_0-Y_0, denotes the *x*-position of point A as x_0, and y-position as y_0. However the point can be addressed differently when other coordinate systems are used. For example, point A can be denoted as x_1, y_1 from the X_1-Y_1 coordinates in Figure 5-4. The coordinates X_1-Y_1 are rotated from the coordinates X_0-Y_0. The relation between the two representations, (x_0, y_0) and (x_1, y_1), can be described as follows:

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \times \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \tag{5.1}$$

where θ is the angle between the two coordinate systems in a clockwise rotation. We can apply this principal to set of scalar values as well. The equation (5.1) can be modified as follows:

$$\begin{bmatrix} \boldsymbol{x}_i \\ \boldsymbol{y}_i \end{bmatrix} = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \times \begin{bmatrix} \boldsymbol{x}_0 \\ \boldsymbol{y}_0 \end{bmatrix} \tag{5.2}$$

where $\boldsymbol{x}_i$=[x_i(1), x_i(2),…, x_i(n)] and $\boldsymbol{y}_i$=[y_i(1), y_i(2),…, y_i(n)]. The scalar values x_i(k) and y_i(k) denotes x- and y-position of the point at time sequence k in X_i–Y_i coordinate system, respectively. The set of scalar values can describe various types of signals including vibration. Thus the vibration signals obtained via proximity sensors as in Figure 5-1 can be regarded as $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$ in equation (5.2).Then, the ODR signals, $\boldsymbol{x}_n$ and $\boldsymbol{y}_n$, can be defined as:

$$\begin{aligned} \boldsymbol{x_n} &= \cos(n\Delta\theta)\,\boldsymbol{x}_0 - sin(n\Delta\theta)\boldsymbol{y}_0 \\ \boldsymbol{y_n} &= \sin(n\Delta\theta)\,\boldsymbol{x}_0 + cos(n\Delta\theta)\boldsymbol{y}_0 \\ & (n = 1, 2, \dots, N) \end{aligned} \tag{5.3}$$

where $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$ are the acquired vibration signals from proximity sensors, $\Delta\theta$ is the increment of the rotation angle, and N (=$\lfloor \pi/\Delta\theta \rfloor$) is the maximum number of ODR that can be generated.

The ODR can generate vibration signals from an arbitrary direction. Multiple ODR signals around the rotor can be obtained by adjusting the increment of the angle, $\Delta\theta$. To diagnose the state of the system accurately, $\Delta\theta$ should be fine. However, if $\Delta\theta$ is too fine, the number of ODR signals (N) will increase, and the computational load will also increase. In addition, the vibration signals are radially symmetric, so ODR signals within the π rotation angle range will be generated. Likewise, there is

no need to use both $\boldsymbol{x}_n$ and $\boldsymbol{y}_n$, because $\boldsymbol{x}_n$ signal is equal to $\boldsymbol{y}_{n+N/2}$, which is 90° rotated signal of $\boldsymbol{y}_n$. The $\boldsymbol{x}_n$ covers all $\boldsymbol{y}_n$ if ODR covers more than half rotation.

5.1.2 Validation of ODR Signals

The ODR signals are transformed from the acquired signals as stated in Section 5.1.1. To use the ODR signals in the diagnosis process, the ODR signals should be verified that those signals are identical to the signals in the same direction. An example of experiment data is used to validate ODR.

First, $\boldsymbol{x}_{N/2}$ which is 90° counter-clockwise (ccw) ODR signal of $\boldsymbol{x}_0$ should exactly match to $\boldsymbol{y}_0$. As shown in Figure 5-5, the direction of $\boldsymbol{x}_{N/2}$ is equal to that of $\boldsymbol{y}_0$, so the two signals should be identical. The two vibration signals in Figure 5-6, prove that they are exactly the same. Second, the direction of $\boldsymbol{y}_{N/2}$ is opposite from the direction of $\boldsymbol{x}_0$, so $\boldsymbol{y}_{N/2}$ and $\boldsymbol{x}_0$ should be symmetric. This is also proved by signals in Figure 5-6. The ODR signal of $\boldsymbol{y}_{N/2}$ is x-symmetric to $\boldsymbol{x}_0$. The last evidence is that the orbit formed by $\boldsymbol{x}_n$ and $\boldsymbol{y}_n$ should be the same. By using both the acquired signals and the ODR signals, orbit shapes are compared in Figure 5-7. Although the vibration signals change over rotation angle, the orbit shape remains constant. From these facts, the ODR signals can be regarded as vibration signals in other directions.

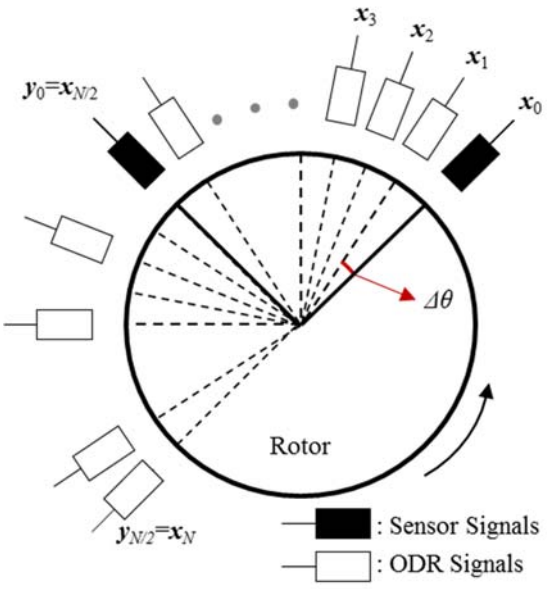


Figure 5-5 Diagram of ODR signals

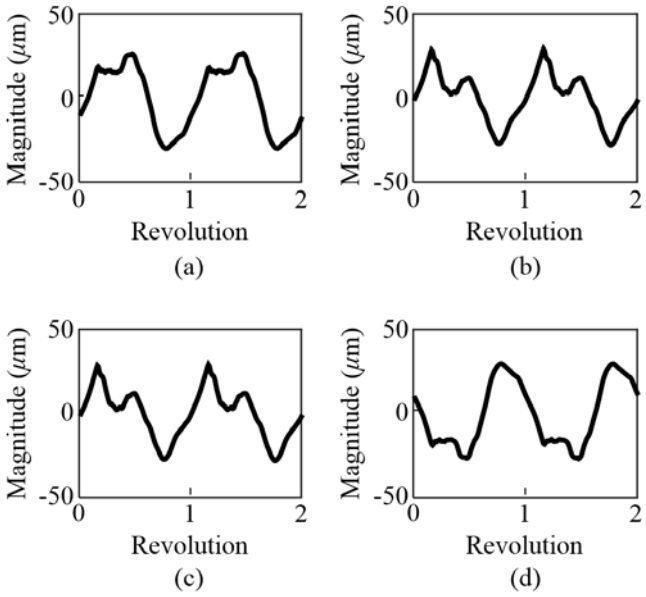


Figure 5-6 Measured signals from (a) $\boldsymbol{x}_0$ and (b) $\boldsymbol{x}_{N/2}$, and ODR signals from (c) $\boldsymbol{y}_0$ and (d) $\boldsymbol{y}_{N/2}$

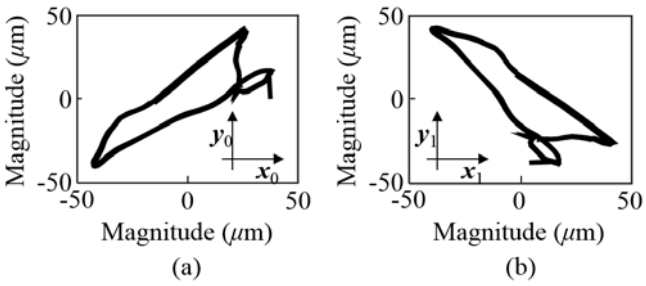


Figure 5-7 Orbits by (a) measured signals ($\boldsymbol{x}_0$, $\boldsymbol{x}_{N/2}$), and (b) ODR signals ($\boldsymbol{y}_0$, $\boldsymbol{y}_{N/2}$)

5.2 Directionality of Health States

Each health state of the rotor system has its own characteristics of vibration signals. Among various health states, few health states show large differences among ODR signals. If only the acquired vibration signals are used for these health states, the vibration signals will vary by the direction of the sensors and anomalies, not by the health states. The inconsistent vibration signals will result in deterioration of the performance of diagnosis. Thus the evaluation of directionality will group health state into either directional or non-directional state. Then the ODR signals can be applied to health state corresponding to the directionality result.

The directional health state denotes that variances exist among ODR signals. An example of an impact-rubbing state signals is presented in Figure 5-8. Among the ODR signals, we cannot predict which signals would be measured through proximity sensors. In some cases, $\boldsymbol{x}_{N/4}$ signal may be obtained, while in other cases, $\boldsymbol{x}_{3N/4}$ signal

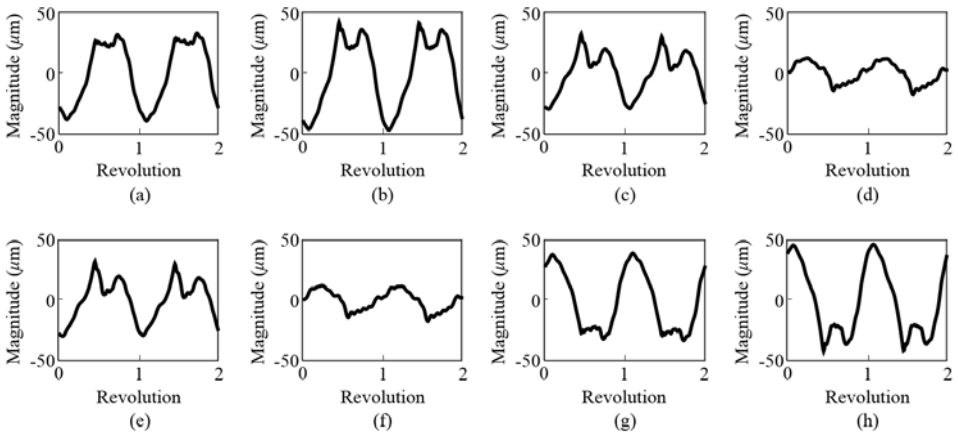


Figure 5-8 Rubbing state ODR signals from sensors (a) x_0, (b) $x_{N/4}$, (c) $x_{2N/4}$,(d) $x_{3N/4}$, (e) y_0, (f) $y_{N/4}$, (g) $y_{2N/4}$, and (h) $y_{3N/4}$

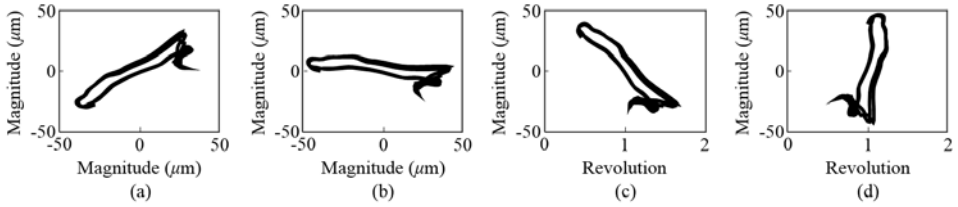


Figure 5-9 Rubbing state orbit from sensors (a) (x_0, y_0), (b) ($x_{N/4}$, $y_{N/4}$), (c) ($x_{2N/4}$, $y_{2N/4}$), and (d) ($x_{3N/4}$, $y_{3N/4}$)

can be obtained. The conventional method which uses only the measured signals would give different diagnosis results for each case, $\boldsymbol{x}_{N/4}$ and $\boldsymbol{x}_{3N/4}$. Thus, for directional health states, the ODR signals should be considered. On the other hand, the non-directional health states have similar vibration signals over rotation angle as presented in Figure 5-10. The waveform of ODR signals for normal state does not change over rotation angle, which indicate that any signal can be used for the

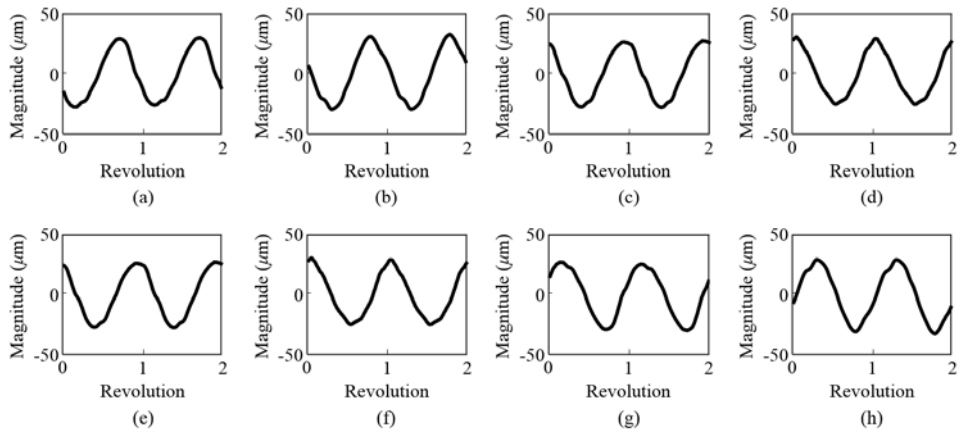


Figure 5-10 Normal state ODR signals from sensors (a) x_0, (b) $x_{N/4}$, (c) $x_{2N/4}$,(d) $x_{3N/4}$, (e) y_0, (f) $y_{N/4}$, (g) $y_{2N/4}$, and (h) $y_{3N/4}$

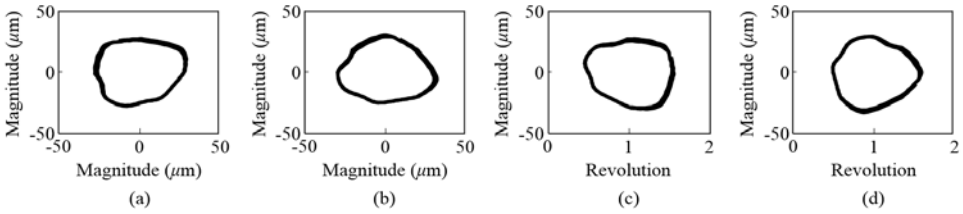


Figure 5-11 Normal state orbit from sensors (a) (x_0, y_0), (b) ($x_{N/4}$, $y_{N/4}$), (c) ($x_{2N/4}$, $y_{2N/4}$), and (d) ($x_{3N/4}$, $y_{3N/4}$)

diagnosis process. The directionality of health state can also be confirmed by the shaft orbit. The orbits of rubbing state in Figure 5-9 express the rotation clearly as different ODR signals are used, whereas those of normal state orbits in Figure 5-11 do not.

The decision of directionality can be made by looking at ODR signals and the orbits, but quantitative metric should be defined for consistent analysis. The metric should work for every health states, and should be irrelevant to anomaly characteristics. So FFT of ODR signals are used. The proposed directionality evaluation metric is defined as follows:

$$D \equiv \frac{\mathrm{Max}\left(S_n(f_{1x})\right) - \mathrm{Min}\left(S_n(f_{1x})\right)}{\mathrm{Min}\left(S_n(f_{1x})\right)} \quad (n = 1,2,\ldots,N) \tag{5.4}$$

where $S_n(f)$ is the power spectrum of the n^{th} ODR signal and f_{1x} is the frequency of rotating speed. The numerator of the metric, D, is the difference between the maximum and the minimum magnitude of power spectrum at rotating frequency (or fundamental frequency) among N ODR signals. Large difference between the two magnitudes indicates that the difference among ODR signals is large, so the health state maybe directional. To make the metric more general, the denominator is added. With the denominator term, D can be compared to that of other systems with different vibration levels. By the empirical study, the health state that has D larger than one is grouped as directional health state, and others as non-directional health state. To validate the metric, D is calculated for the rubbing and normal states. The result in Figure 5-12 shows that D of rubbing exceeds ones, while D of normal does not.

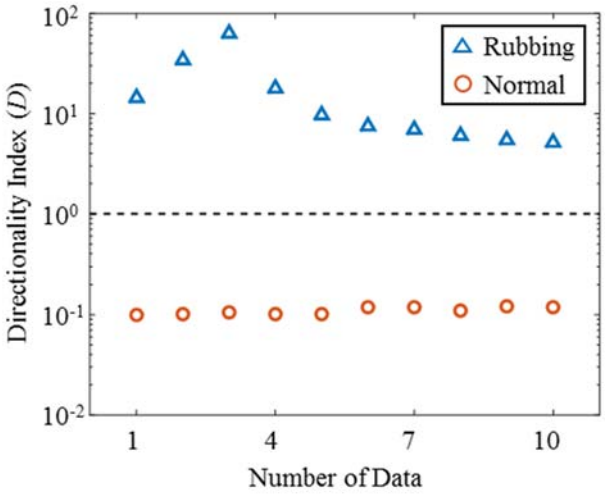


Figure 5-12 Example of Direction Evaluation Result

5.3 Health Classification using ODR Signals

As stated in Section 5.2, the rotor systems can be grouped into either directional or non-directional health state. To determine the directionality, ODR signals are generated for all health states. With the ODR signals, the directionality are obtained through evaluation metric proposed in Section 5.2. The overall health classification procedures based on directionality of ODR signals are presented in Figure 5-13. For the directional health states, we should use all N ODR signals for the robust diagnosis. The features are extracted and selected from each ODR vibration signal, and the selected optimal features are used for the classification process. The k selected

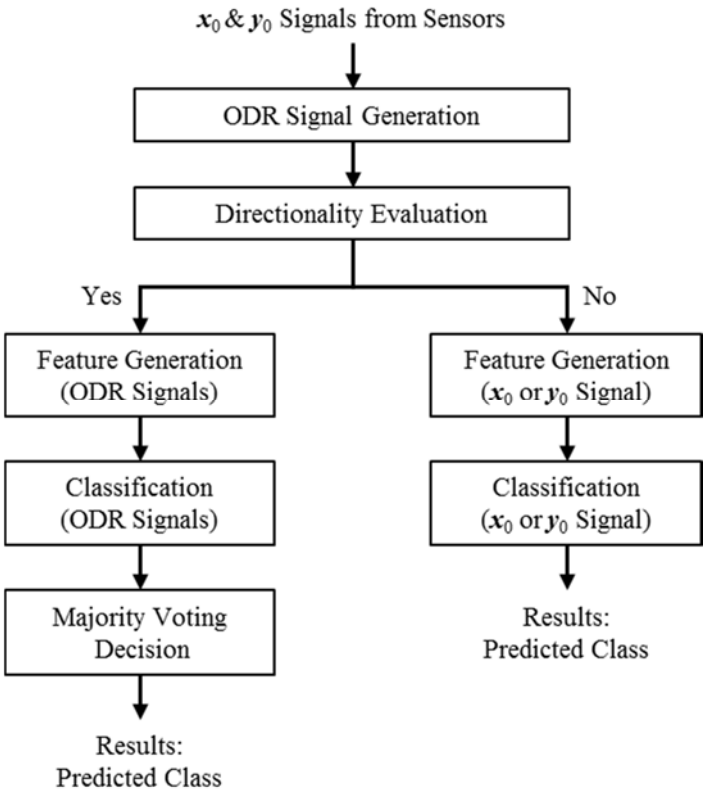


Figure 5-13 Overall procedures of ODR based diagnosis

features of each *N* ODR vibration signal are used for training the classifier. The support vector machine algorithm trains the classifier with the labeled data. Then, the class of unlabeled data of each *N* ODR vibration signal is predicted. The prediction gives *N* results for each ODR vibration signal. The *N* predicted classes may have the same results, however if not, a decision should be made to determine the class of the unlabeled data.

In this research, majority voting scheme is used for the final decision. Originally, the majority voting scheme is frequently used to integrate the results from multiple classifiers. For example, each classification result of support vector machine (SVM), Fisher's discriminant analysis (FDA), and artificial neural network (ANN) are obtained for the same data. The final decision can be made by selecting the most frequent classification result. Benchmarking the majority voting scheme, we have applied this scheme to the *N* results of ODR signals. Among the *N* predicted classes, the majority class will be the final prediction result. For cases with multiple majority classes, the prediction result will give no result, which indicates that the prediction result is not accurate. An example is shown in Figure 5-14. The majority voting scheme is used for the final prediction of sixteen ODR signals of rubbing state. As presented in Table 5-1, the majority of sixteen classes is the rubbing health state

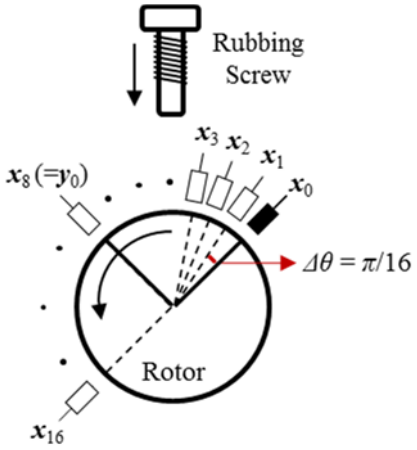


Figure 5-14 ODR signals at rubbing state

Table 5-1 Example of predicted class using ODR signals

ODR	Predicted Class	ODR	Predicted Class
x_1	Rubbing	x_9	Rubbing
x_2	Rubbing	x_{10}	Rubbing with Unbance
x_3	Rubbing	x_{11}	Misalignment
x_4	Rubbing	x_{12}	Misalignment
x_5	Rubbing	x_{13}	Rubbing
x_6	Rubbing	x_{14}	Rubbing
x_7	Rubbing	x_{15}	Misalignment
x_8	Rubbing	x_{16}	Rubbing with Unbance

which counts for eleven ODR signals, thus the final prediction of the example is the rubbing state.

The majority voting results can be changed by the number N. So the optimal number for N should be determined by the diagnosis accuracy. The definition of N ($=\lfloor \pi/\Delta\theta \rfloor$) is inversely proportional to the increment of the rotation angle, $\Delta\theta$. If the

$\Delta\theta$ is small, the fineness of ODR signals is increased, which gives better chance of understanding the health state. However, too small ODR signals will cost much computational load without improving the performance or it will yield overfitting problem. Thus N should be compromised or optimized between the prediction accuracy and the computational load.

5.4 Results of ODR

5.4.1 ODR Signals for Health States

The time and FFT plots can display the physical interpretation of each health state. This section provides the time and FFT plots of the ODR signals as well as measured signals of the test data. Total of five health states—normal, rubbing, misalignment, oil whirl, rubbing with unbalance—were tested. From each test, two vibration signals ($\boldsymbol{x}_0$ & $\boldsymbol{y}_0$) were measured by proximity sensors in orthogonal direction. Based on $\boldsymbol{x}_0$ & $\boldsymbol{y}_0$, ODR signals were generated as in Figure 5-15. In addition FFT plots for each ODR signal are presented in Figure 5-16 for further understanding.

The signals of normal state is close to the typical sinusoidal wave irrespective of ODR rotation. This is also confirmed in the FFT plot that 1x frequency component is dominant for all ODR signals. Next, the oil whirl state is distinguished by the FFT plots of 0.42-0.48x dominant case [49, 130, 131]. The oil whirl is induced by

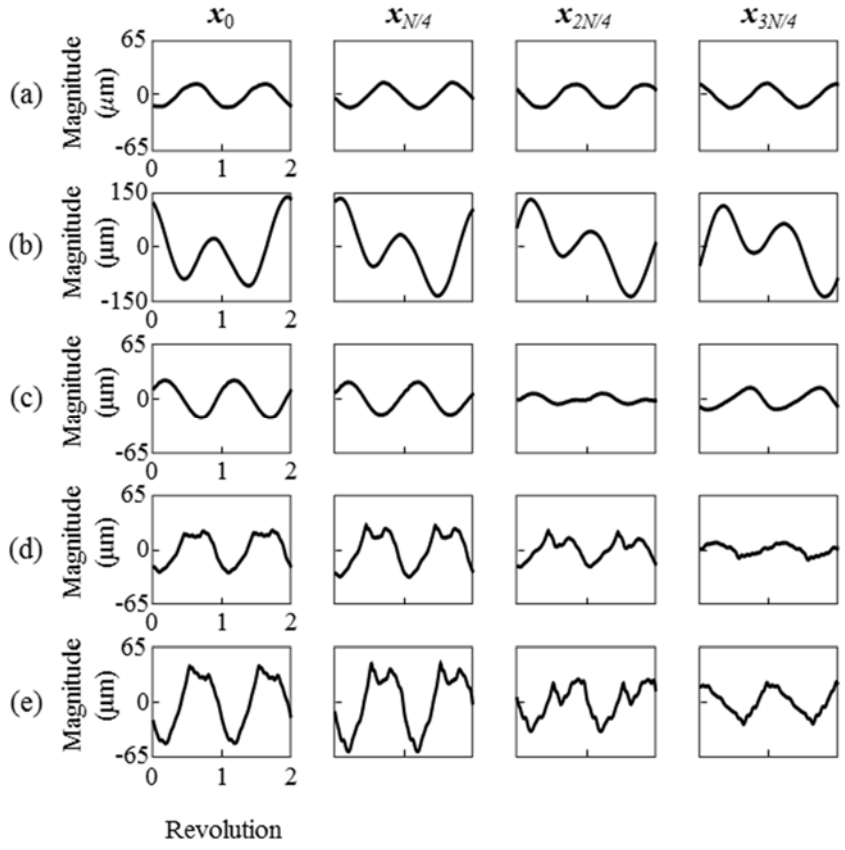


Figure 5-15 ODR signals of health states (a) normal, (b) oil whirl. (c) misalignment, (d) rubbing, (e) rubbing with unbalance

instability of lubricant oil surrounding the shaft. Thus the characteristics are similar over the rotation angle as presented in time plots and FFT plots. For misalignment health state, some of the time plots resemble that of the normal case, but time plots change over the ODR rotation angle. Especially, some ODR signals have very small amplitude with 2x frequency component, which is the result of misaligned shaft. For the rubbing health state, time plots change greatly as the ODR rotation is changed.

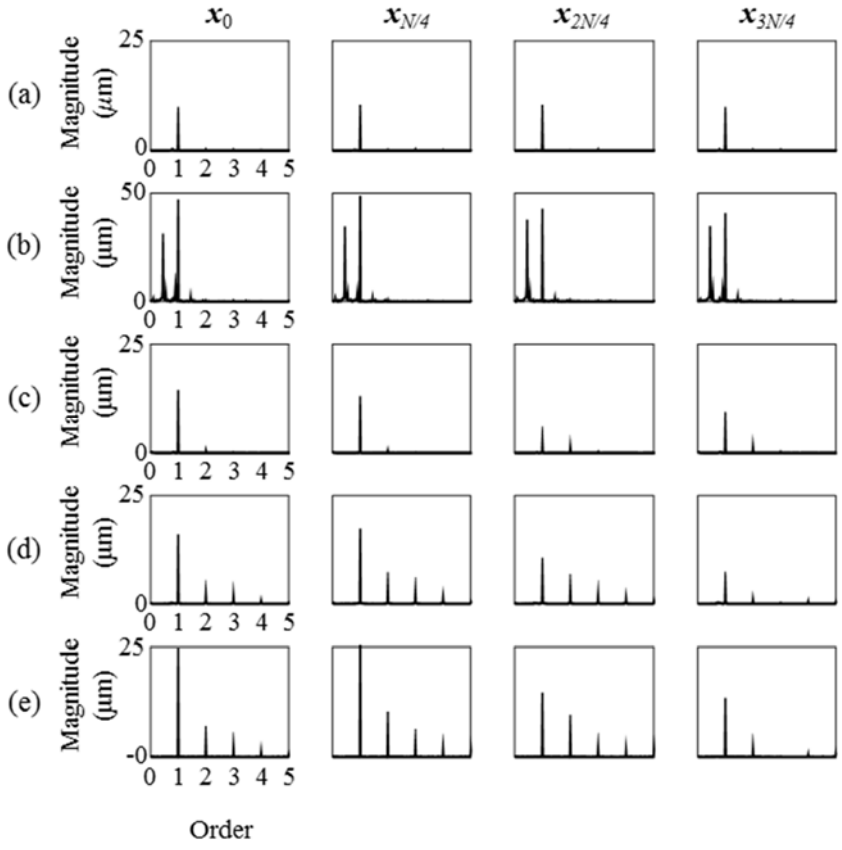


Figure 5-16 FFT of ODR signals corresponding to Figure 5-15

In some directions, the trimmed sign of sinusoidal signal due to impact-rubbing is obvious. But in some other directions, the signals are very hard to tell whether impact-rubbing is made. The FFT plots of rubbing also differs over the ODR rotation angle. Some FFT plots have 1x dominant with harmonic components, which is a typical rubbing sign. Other plots have very small 1x frequency which makes other frequency components dominant relatively. The rubbing with unbalance case is

similar to the rubbing case. The amplitude of 1x component in FFT has been increased and this is revealed in time plots as well.

From the figures of FFT, normal and oil whirl health states can be grouped as non-directional states. The time plot of oil whirl seems to change over rotation angle, but the variance of time and frequency features are minor compared to those of

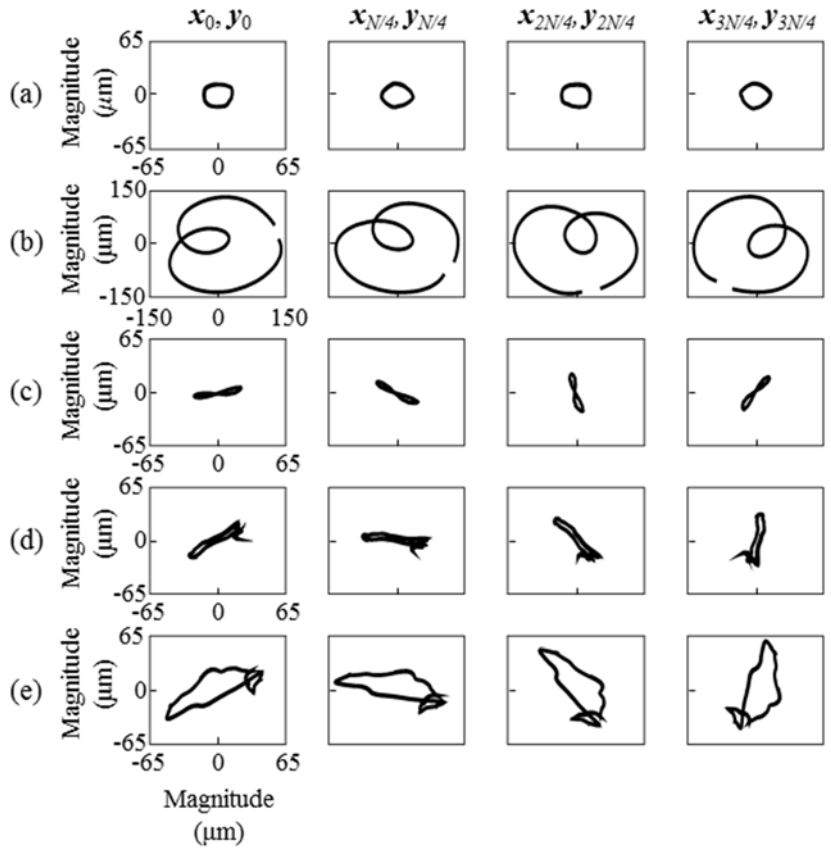


Figure 5-17 Orbit of ODR signals corresponding to Figure 5-15

directional features. An example of distribution of time feature is presented in Figure 5-18. The figure shows that the variance of oil whirl features are much smaller than those of rubbing features. This can also be confirmed by looking at the FFT plots. The FFT plots of normal and oil whirl states do not change much over rotation angle, while FFT plots of other states change greatly. Thus we know that ODR signals should be applied to rubbing, rubbing with unbalance, and misalignment cases for robust characterization of each health state. However, for consistent analysis of the directionality, the metric that is stated in Section 5.2 is applied in the next section.

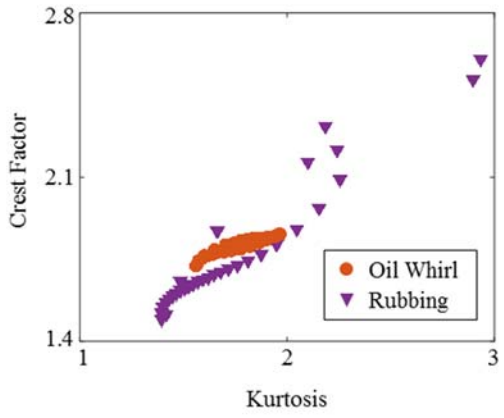


Figure 5-18 Distribution of Kurtosis and Crest Factor values for oil whirl and rubbing states

5.4.2 Directionality for Health States

The directionality of health states is determined by the directionality evaluation metric, *D*. The objective of evaluating the directionality is to split the health data into two groups, directional or non-directional states. For directional states, ODR signals are to be applied for robust characterization, whereas for non-directional states, ODR signals are redundant of acquired signals. All the data used in this research is evaluated for directionality. A set of data includes sixty seconds signals of each health state. Three sets in total were evaluated and the result for each validation case is presented in Figure 5-19.

For data set 1, the normal and the oil whirl states had *D* values less than one, which indicates non-directional health state. The other health states showed *D* values bigger than one, which denotes directional health state. The other two data sets had consistent results with the first case. Hence the first two health states can be grouped

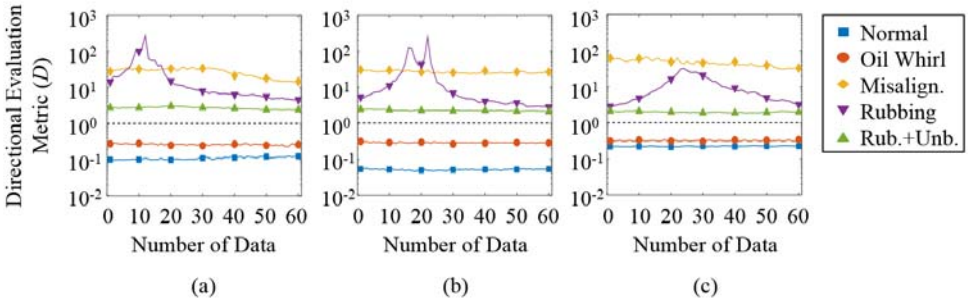


Figure 5-19 Results of directional evaluation metric for (a) data set 1, (b) data set 2, and (c) data set 3

as non-directionality group, whereas the last three health states can be grouped as directional group.

For rubbing and rubbing with unbalance states, a point contact occurs in a vertical direction. The effect of the contact is not seen in all directions but in a certain direction. The most evident trace of the contact is represented when the direction of ODR matches to that of contact. Thus, health states including rubbing can be defined as directional states. For misalignment state, the shaft is shifted in a horizontal direction, which leads to increase of horizontal stiffness on the bearing. This leads to uneven magnitude of the response between horizontal and vertical signals, which causes the ODR signals to change over rotation angle.

5.4.3 Classification Results by ODR

The five health states are grouped into either directional or non-directional state. For directional state group, the features are extracted from all ODR signals. Using the extracted features, the training of the classifier and the prediction of the unlabeled data are performed. Since N results are obtained for each sixty-cycle data, the final prediction result is obtained by majority voting. For non-directional state group, the features are extracted from one of the measured signals. The ODR signals are not used for this group because time and frequency features do not change much

compared to those of directional states. With the extracted features, the training and prediction processes are performed.

To validate the effectiveness of ODR signals on directional health states, the class prediction result is presented. The two results including with and without ODR signals are compared. In addition, the results of the non-directional health state is also presented.

ODR vs. Non-ODR

First, the effectiveness of the ODR method can be validated by the results in Figure 5-20 for directional health states. The class prediction accuracy of ODR and non-ODR signals are compared in the figure. The lines marked by circles represent

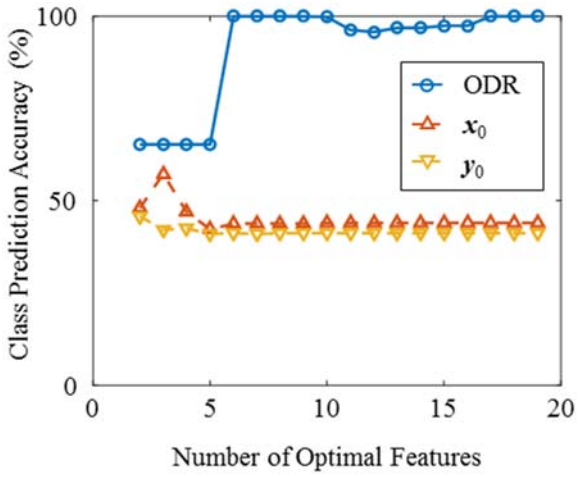


Figure 5-20 Classification results of directional health states using measured signals

the results by the proposed ODR methods, while the lines marked by upward- and downward-triangle represent the results by non-ODR signals. For both cases, the optimal number of features (n) for feature selection process are increased from two to nineteen. Each line represents average of three cross validation cases by three data sets. The results by ODR signals are based on the majority voting scores of sixteen ODR signals ($\Delta\theta = 11.25°$). Contrast to that, other lines are the results of non-ODR signals, which uses the measured signals, $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$, for training and prediction.

The classification results clearly show that the ODR method outperform the conventional method which uses $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$ signals separately. When six or more features are used, the ODR method classifies the given data without many misclassifications. However, the results by $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$ signals show less than fifty-percent for most of the feature numbers. These results indicate that $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$ cannot fully characterize the directional health states. In addition, the class prediction results using $\boldsymbol{x}_n$ and $\boldsymbol{y}_n$ signals are presented in Figure 5-21. To consider the uncertainty of directions, all possible combinations of signals for prediction were used with $\boldsymbol{x}_0$ and $\boldsymbol{y}_0$ training. Since N signals were generated by ODR, N^3 combinations for three health states were used for the prediction. The average of N^3 results are presented as lines marked by triangles.

As expected, the prediction accuracies of the ODR method are substantially higher than those of the non-ODR method. Most of the predictions by the ODR method reach accuracy of hundred percent when number of optimal features were

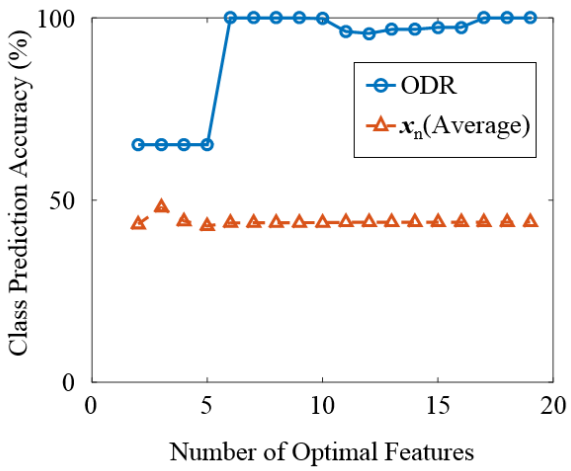


Figure 5-21 Classification results of directional health states: Average of all feasible combinations

larger than five, while the average of the non-ODR method remain less than fifty percent irrespective of optimal features. However, the ODR method of five or less optimal features show about seventy percent due to misclassification of rubbing with unbalance case. The misclassification is reduced after amplitude related features are included in the optimal features. In addition, the low accuracies of the non-ODR combinations had accuracies between zero to hundred percent, which means that prediction heavily depends on the direction of abnormality and sensors.

Second, the class prediction results for non-directional health states are shown in Figure 5-22. Likewise, the class prediction accuracy of ODR and non-ODR are presented in the figure. Similar to the directional case, the ODR method uses *N* ODR signals for training and prediction, whereas the non-ODR method uses only one

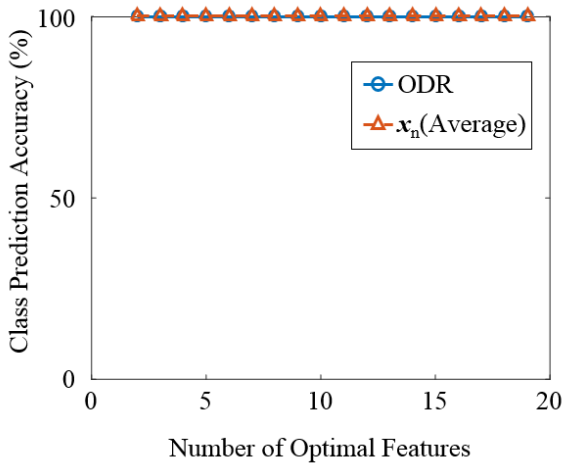


Figure 5-22 Classification results of non-directional health states: Average of all feasible combinations

signal. The non-ODR results are the average prediction accuracy of N^3 combinations. The result indicates that both ODR method and non-ODR method are valid for classification of non-directional health states. Not much difference exists between the ODR method and the non-ODR method because signals do not vary with respect to direction. Thus, if the health states are grouped as non-directional by evaluation metric, a signal from any direction can be used for the training and prediction processes.

Chapter 6. Unsupervised Learning Framework for Power Plant

We introduced supervised learning for journal bearing rotor system using test-bed data with consideration of physics. However, the application of developed algorithm in small-scale test-bed to the diagnosis of actual system such as rotors of a large-scale power plant is needed additional consideration for the un-tested conditions. Unsupervised learning techniques can help this problem. For these purpose, we would like to suggest the diagnostic framework for a power plant using deep learning architecture which has high performance for supervised and unsupervised recognition using high-level feature abstraction.

6.1 Overview of Deep Learning for Diagnosis

Resources of diagnosis algorithm development for Journal bearing rotor systems are classified as labeled data (for supervised learning) and unlabeled data (for unsupervised training). Detail categories for acquired signals from the test-bed and a power plant are shown in Figure 6-1. The labeled data can be used for validation of supervised learning directly by using classification prediction accuracies whereas it can also be used for validation of unsupervised learning indirectly by the comparison of the predicted cluster of known conditions.

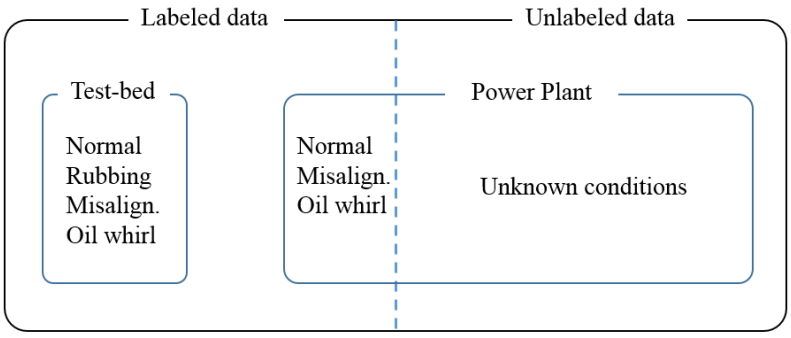


Figure 6-1 Data categories of journal bearing rotor systems

The goal of this chapter is execution of unsupervised learning for a power plant through reasonable validation process. In order to develop unsupervised training for a vibration signals of power plant, some validation stages are required. We established three stage developing procedures from test-bed to a power plant case as shown in Figure 6-2. Already we know the success of deep learning in an image recognition. Therefore, the image type of vibration data which contains overall characteristics of rotor behavior were used for input signal of DBN. In this study, five layers of hidden stage for DBN structures was used for high level feature abstraction as shown in Figure 6-2. The details of DBN and RBM will be explained in the Section 6.2.

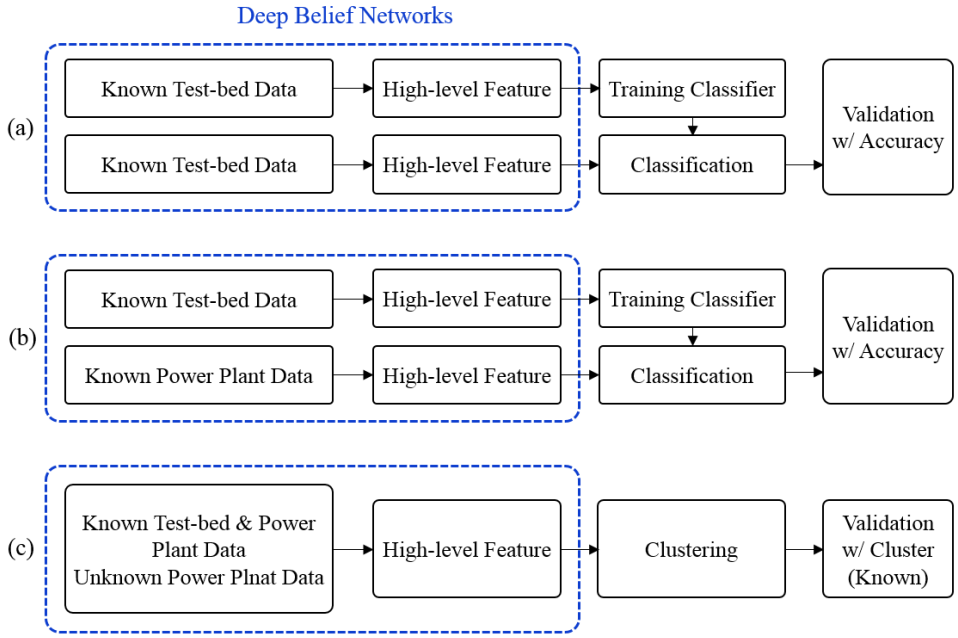


Figure 6-2 Diagnosis procedures for (a) classification and validation of test-bed data (b) classification and validation of power plant data (c) clustering of test-bed and power plant data

6.2 Deep Learning Architecture of Gap Sensor Signals

The vibration data for image generation are based on the ODR signals. These signal contains overall vibration characteristics of rotor system as well as gives intuitive understanding of health states.

6.2.1 Image Generation for Deep Learning

The vibration data measured by the gap sensors are pre-processed through the process in the Figure 6-3 for deep learning. It consists of resampling with low-pass filter, phase synchronizing, stacking with ODR signal generation, and normalization. Resampling and low-pass filter were mentioned in section 3.1. Especially, low-pass filter is efficiently effected for the reduction of high frequency noise of the actual power plant data which will contribute robust procedure for diagnosis. After resampling the gap sensor signals are synchronized in time-domain. The physical meaning of gap sensor signals in the image domain are differ according to its phase information. The results of image generation in various phase cases for the same

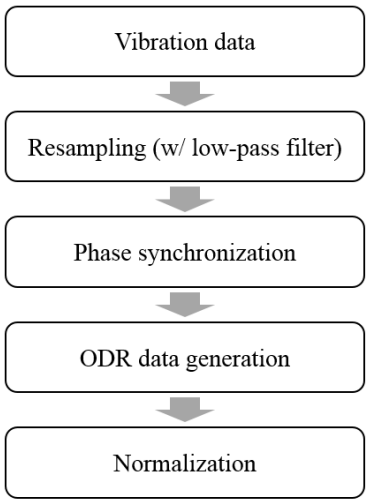


Figure 6-3 Pre-processing of gap sensor signals for deep learning

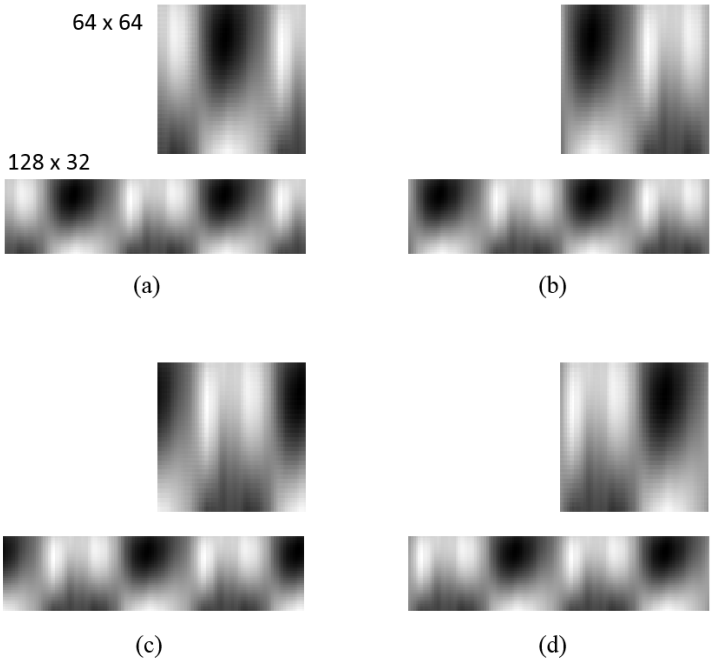


Figure 6-4 Examples of ODR imaging (64x64 or 32x128 pixels) in the cases of phase delay from 1st peak of 1x signal in (a) 0 (b) π/2 (c) π (d) 3π/2

health state are shown in Figure 6-4. The images are shown in the size of 64x64 and 32x128 pixels according to the methods stacking ODR signals. The vibration images are absolutely dependent to the phase of the signals.

Phase Synchronization

Phase synchronization offers the same criteria in sequence for feature-domain to make similar images of gap sensor signals in the similar behavior. In most case, 1x signal which is harmonic component of fundamental frequency is dominant in

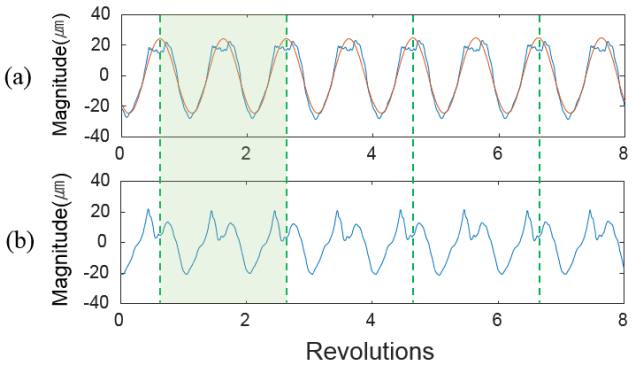


Figure 6-5 Gap sensor signals for eight revolution in (a) x-direction and (b) y-direction

journal bearing rotor system. Therefore 1x signal can be used for time criteria of phase synchronization. In this work, the first peak of 1x signal are used as the starting point of signal generation. Figure 6-5 shows gap sensor signals for eight revolution in x- and y-direction. Total vibration and 1x signals are represented in blue and red line, respectively. For each two cycle of gap sensor signals (green colored region in Figure 6-5) are used for image generation. In the vibration signals acquired in 3,600 RPM, thirty vibration images could be generated from the vibration data for sixty seconds. The generated vibration images are the outputs of ODR signals, which will be explained in the followings.

Vibration Images from ODR Signals

ODR described in Section 5 are formulate the high diagnostic performance because it represents the overall behavior of the system status. Applying the ODR

technique in the journal bearing system can create the image reflecting the overall omnidirectional vibration characteristics. This ODR based vibration image enables the robust fault diagnosis regardless the orientation of measurement and the direction of anomaly occurrence.

The generation of vibration image come from stacking ODR signals. Figure 6-6(a) shows the ODR signal mentioned in Section 5.1. The positions of x_0 and y_0 in Figure 6-6(a) mean the directions of gap sensor installation. Phase synchronized x_0 and y_0

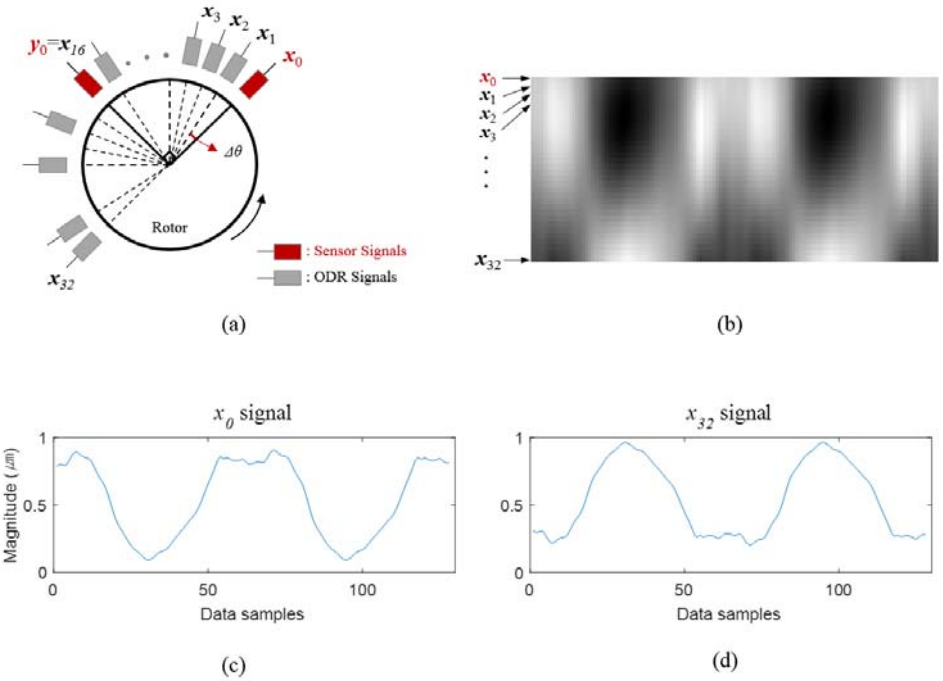


Figure 6-6 (a) Diagram of ODR signals (b) vibration image using ODR signals with normalization (c) ODR signal in x_0 (d) ODR signal in x_{32}

signals for two revolutions are used for ODR generation, which are resampled signals in 64 samples per each cycle. Two revolutions as a length of data for image generation are employed to improve the effectiveness of vibration characteristics representation. The ODR signals are generated in 32 directions which are represented in the direction of gray colored sensor position. Vibration image from the stacked ODR signals is shown in Figure 6-6(b). ODR signals are normalized 0 to 1 by using the maximum and minimum of total ODR data in two revolution. The vibration image displays pixels with the value 0 as black and 1 as white. The examples of normalized ODR signals in the x_0 and x_{32} direction are shown in Figure 6-6(c) and (d). The vibration images for test-bed data in the states of normal, rubbing, misalignment and oil whirl are presented in Figure 6-7.

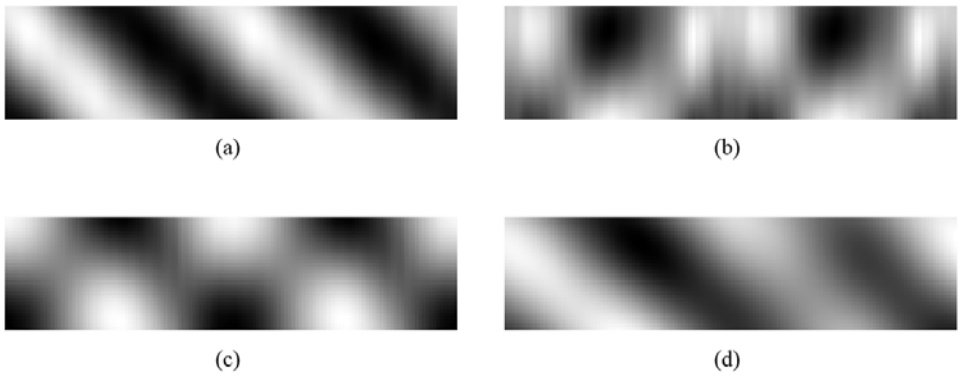


Figure 6-7 Vibration images for test-bed data in the condition of (a) normal (b) rubbing (c) misalignment (d) oil whirl

6.2.2 Generation of High-level Features

Vibration images generated from ODR signals introduced in Section 6.2.1 are used as inputs of DBN. The description of data used in the deep learning study is presented in Table 6-1. Labeled data in test-bed and power plant which are used for the quantification of reasoning performance consist of four health conditions such as normal, rubbing, misalignment, and oil whirl. The input vibration image is 4096 (32x128) dimensional data. Candidate sizes of hidden nodes for stacked RBM were used 2048, 1024 and 512 dimensions which are kind of divisor for the dimension of input vibration image. Then, three to five hidden layers were employed for possible RBMs in deep learning. The number of hidden layer and the size of hidden nodes

Table 6-1 Description of data used for deep learning

System	Status	Health condition	Data size of vibration image (Number x Dimension)
Test-bed	Labeled	Normal	1000 x 4096
		Rubbing	1000 x 4096
		Misalign.	1000 x 4096
		Oil whirl	1000 x 4096
Power Plant	Labeled	Normal	1000 x 4096
		Misalign.	1000 x 4096
		Oil whirl	1000 x 4096
	Unlabeled	-	15000 x 4096

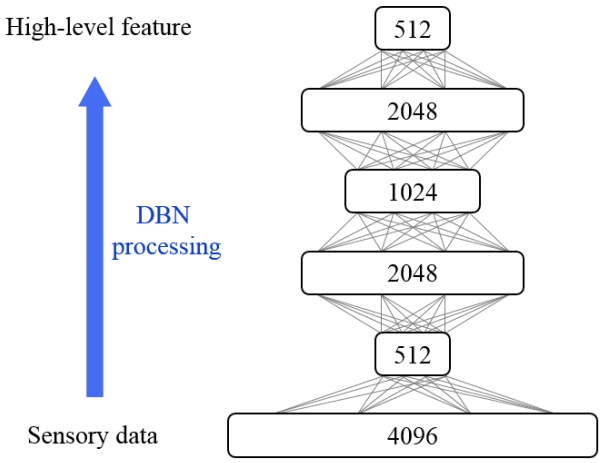


Figure 6-8 DBN using five hidden layers of which sizes are 512, 2048, 1024, 2048, and 512 dimensions

are selected from the classification result of labeled data in Table 6-1. In this study, deep learning employed five stacked RBMs whose dimensions for each hidden layer are 512, 2048, 1024, 2048, 512, respectively, as shown in Figure 6-8. The classification results for supervised data using given DBN are represented in Section 6.3.1.

6.2.3 Reasoning Algorithms

Deep learning algorithms are used for two kinds of diagnostic methodologies: health state classification and clustering. An appropriate classifiers are required according to the diagnosis algorithms. We selected the classifiers from the branch of

ANNs. The multilayer perceptron (MLP) and self-organizing map (SOM) employed as classifiers for classification and clustering, respectively.

Multilayer Perceptron (MLP) for Classification

Multilayer perceptron (MLP) is a type of feedforward neural networks (FFNNs) that mapping multi-dimensional inputs onto newly generated outputs [132, 133]. MLP basically consists of three layers: input layer, hidden layer, output layer. A hidden layer can be designed with a single layer or multiple layers of neurons, with each nodes of a layer fully connected to the nodes of next layer. The weighted sum of each layer and bias temp processed by a nonlinear activation function for generating node of next layer or outputs. MLP using a backpropagation technique is one of the most popular algorithm for any supervised learningpattern recognition process.

Self-organizing Map (SOM) for Clustering

The SOM is structured by nodes in a one- or two-dimensional regular grid, which are configured as hexagonal or rectangular configuration [134, 135]. A schematics of SOM with 2-D hexagonal lattice are illustrated in Figure 6-9. The SOM essentially defines a mapping from the input ($\mathbf{x} \in \boldsymbol{R}^n$) onto a 2-D array of neurons. Every neuron of SOM is associated with an n-dimensional weighting vector $\mathbf{w}_i$ where $\mathbf{w}_i = [w_{i1}, w_{i2}, w_{i3}, \dots w_{im}]^T$, m is the dimension of the input data. An m-dimensional input vector ($\mathbf{x} = [x_1, x_2, x_3, \dots x_m]^T$) is connected to all component of weighting

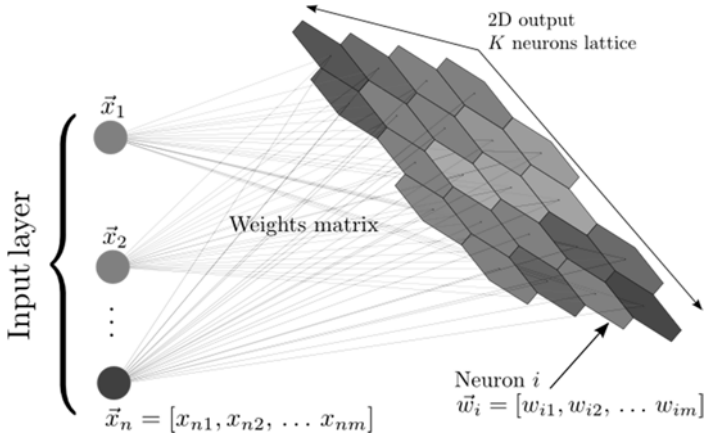


Figure 6-9 A schematic representation of a SOM [135]

vectors. Additionally each neuron can be affected by neighbor neurons. Once all weighting vectors of neurons have been initialized, the Euclidean distances between and input and all weighting vectors are calculated. The weighting vector which is smallest Euclidean distance is called as best-matching unit (BMU). The equation for searching BMU is denoted as:

$$c = \arg\min_i(\|\mathbf{x} - \mathbf{w}_i\|) \tag{6.1}$$

All weighting vectors are updated according to the following equation as:

$$w_i(t+1) = w_i(t) + h_{ci}(t)[x(t) - w_i(t)] \tag{6.2}$$

where $h_{ci}(t)$ is the neighborhood kernel around the BMU *c* at the time-step *t* in updating process of SOM. The neighborhood kernel is defined in the following equation as:

$$h_{ci}(t) = \alpha(t) \cdot h(\|r_c - r_i\|, t) \tag{6.3}$$

where r_c and r_i are the location vectors of neurons *c* and *i*, $\alpha(t)$ is learning rate in the range of 0 to 1. For the convergence of the SOM process, $h_{ci}(t)$ goes to 0 when $t \to \infty$, and with increasing $\|r_c - r_i\|$.

6.3 Results of Deep Learning

The diagnosis results using deep learning consist of the supervised and unsupervised case. Data obtained from test-bed and known data of power plants were performed a supervised learning. Then the whole data including unknown status of power plants were performed unsupervised learning.

6.3.1 Supervised Learning Results

Health Classification of Test-bed Data

Experiments with test-bed were performed for the four conditions of health states – normal, rubbing, misalignment, and oil whirl. The experiments are carried out five

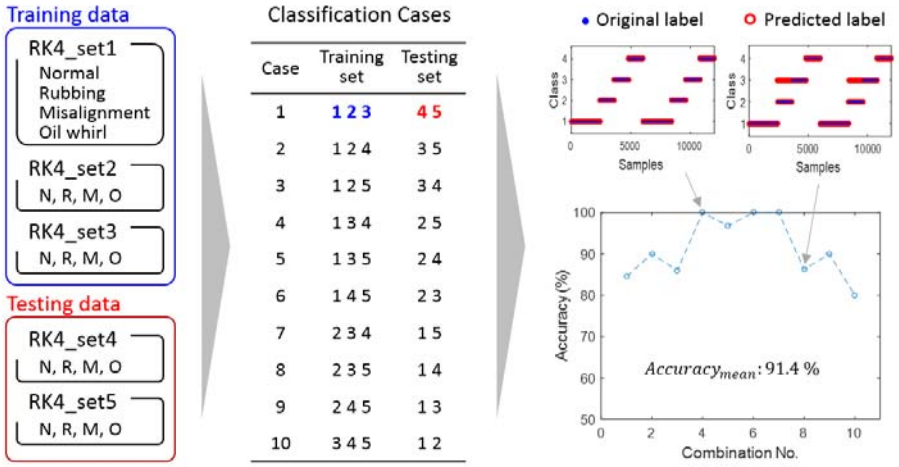


Figure 6-10 Procedures for classification of test-bed data (a) five sets of test-bed data in four anomaly conditions (b) the combinations of training and testing data set (c) classification accuracy and its meanings

times for each health condition, and each experiment represented by set as show in Figure 6-10(a). To obtain classification results, three sets of experimental data were used as training data, and the rest two sets were used as testing data. The performance of diagnosis using health classification was obtained by the averaging of results for all the possible 10 combinations as shown Figure 6-10(c).

In this study, we examined four case of classification results in order to validate the effect of deep learning, and MLP with 30 hidden neuron was used as classifier. Data used for four classification cases are as follows: raw vibration data, raw data

with deep learning, ODR data, and ODR data with deep learning. The performance of health classification for the four cases mentioned above are shown in Figure 6-11. Class 1 to 4 in Figure 6-11 are matched to the condition in normal, rubbing, misalignment, and oil whirl, respectively. When used raw data, there is not much difference in diagnostic performance regardless the implementation of deep learning and indicates relatively low accuracies as shown in Figure 6-11(a) and (b). On the other hand, if you use the ODR data, diagnostic performance are highly improved.

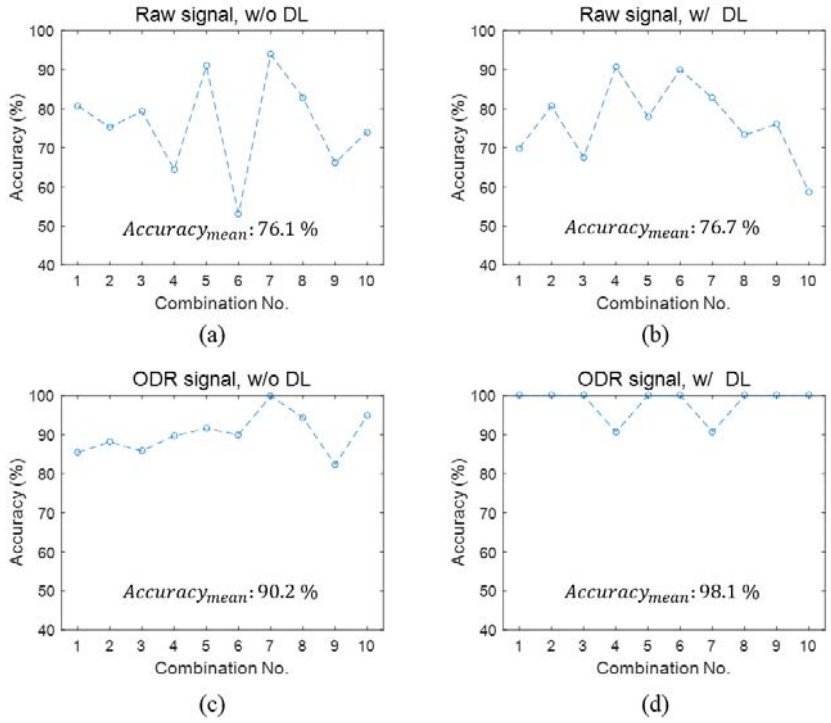


Figure 6-11 Classification prediction accuracy according to selected data and algorithms (a) raw vibration signals without deep learning (b) raw vibration signals with DL (c) ODR signals without DL (d) ODR signals with DL

The classification result using the ODR with deep learning presents the best diagnostic performance as shown in Figure 6-11. Although some samples in rubbing condition in set 4 and 7 in Figure 6-11(d) are misclassified, it shows high classification accuracy as 98.1%. These results represent that the high-level feature generated by deep learning shows good performance in supervised learning.

Health Classification of Power Plant Data

In order to investigate the health classification performance of the power plant data, high-level features from test-bed and known condition of power plant described in Table 6-1 were used as training and testing data. Power plant data contains three health conditions: normal, misalignment, and oil whirl. Class l through 4 in Figure 6-12 means normal, rubbing, misalignment, and oil whirl, respectively. Although

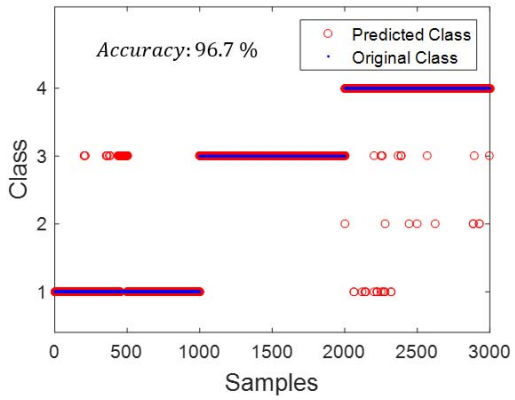


Figure 6-12 Classification prediction results from training with test-bed data and testing with labeled power plant data

few misclassified results in normal and oil whirl are exist, the classification results confirmed high classification accuracy as 96.7 %. These result shows high potentials of high-level features driven by deep learning for power plant diagnosis.

6.3.2 Unsupervised Learning Results

Unsupervised learning trains the algorithm without the information of label. If we have some labeled data, it can be used for the validation of clustered results or as a basis of semi-supervised learning. Unsupervised learning used total 22,000 number of data described in Table 6-1 which are acquired from test-bed and power plants in the conditions of known and unknown. Each data for learning is high –level features extracted from deep learning in 512 dimension. Among all the data for unsupervised learning, the conditions of 32% data are known and the rest are unknowns as illustrated in Figure 6-13. SOM classifier with 9 cluster was employed as the

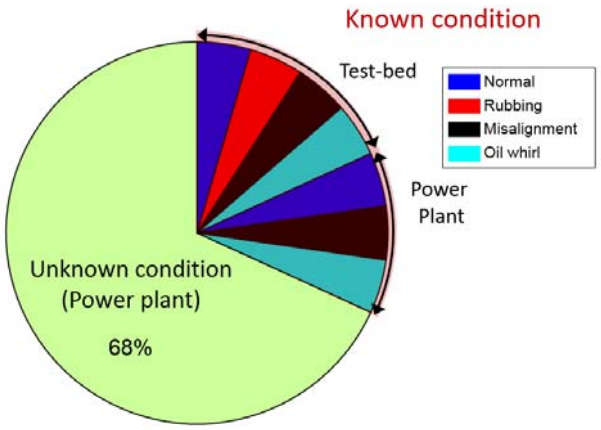


Figure 6-13 Data composition for unsupervised learning

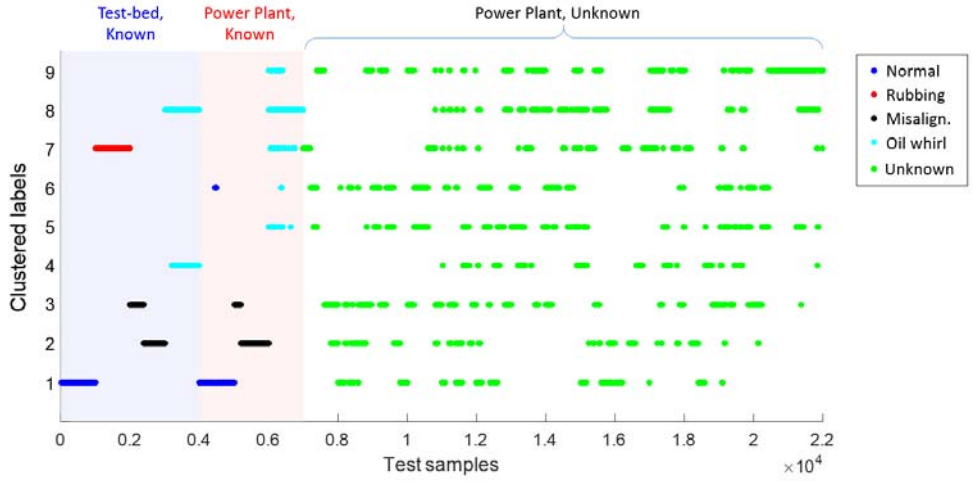


Figure 6-14 Results of unsupervised clustering with high-level features by deep learning

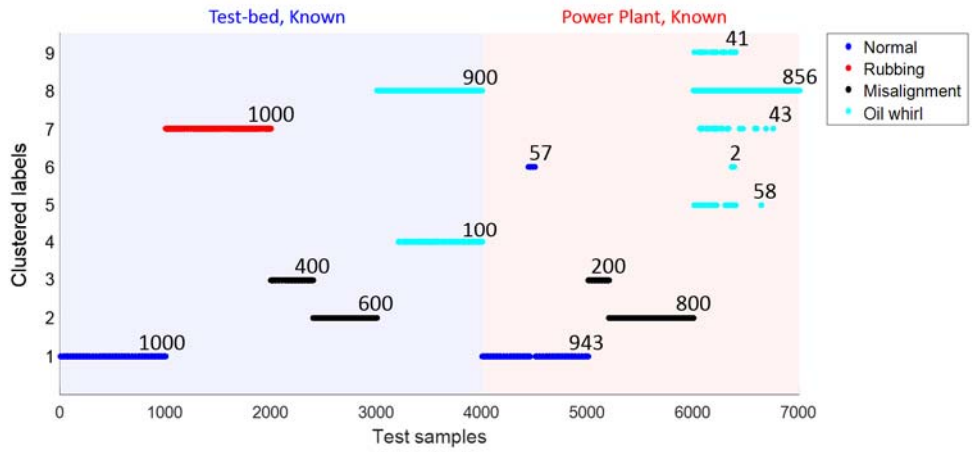


Figure 6-15 Clustering results of supervised conditions in test-bed and power plant

classifier for clustering. The results of unsupervised learning are shown in Figure 6-14. Zone in light-blue and light-red in Figure 6-14 represent the results of test-bed and known conditions in power plants. The specified results for known condition

are shown in Figure 6-15. Clustered results of known power plants are well matched to the results of test-bed as 94.7% accuracy. We obtained the accuracy by dividing the numbers which resulted same class in test-bed and known power plant by total number of known power plant data. The occurrence of each clustered label for total testing data are presented in. Figure 6-16. The cluster 1 and 9 represents high occurrences than others. Cluster 1 seems to normal condition, whereas cluster 9 remained unknown conditions. Cluster 5 and 6 are also remained unknown status in unsupervised results. The clustering results of unknown conditions are evenly distributed as shown in Figure 6-16(b). This is because the unknown data of the power were obtained in the condition of the anomalies intermittently generated. So the occurrence of several evenly distributed clustering results are reasonable in

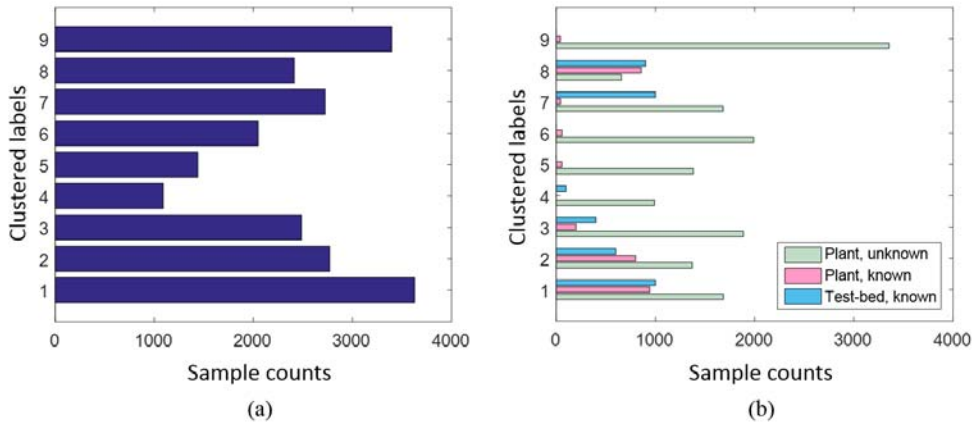


Figure 6-16 Histogram of clustered results using high-level features by deep learning

unstable power plant conditions. Figure 6 17 represents a kind of validation for clustered results whose vibration images for unknowns in power plant are similar to vibration images in test-bed for the case of same clustered label.

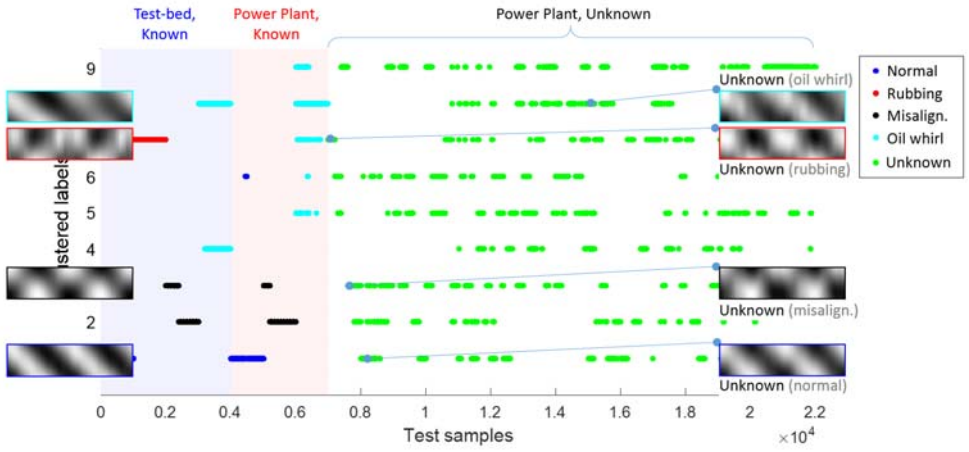


Figure 6-17 Validation of clustered for unsupervised status by vibration images

Chapter 7. Contributions and Future Works

7.1 Contributions and Impacts

The proposed research in this thesis aims at developing robust diagnosis algorithm for journal bearing system in a power plant and to advance conventional diagnosis techniques using gap sensors. This research is composed of three research thrusts: 1) datum unit optimization, 2) omnidirectional regeneration of gap sensor signals, and 3) unsupervised learning framework for a power plant. It is expected that the proposed offers the following potential contributions and broader impacts in various engineering fields:

Contribution 1: Suggestion of optimal datum unit for diagnosis using gap sensor signals

This research defines the optimal datum unit for the diagnosis of the journal bearing system. From the classification study for the anomaly diagnosis, it can be concluded that the proposed optimal datum unit guarantees the high separability. Since the high separability enables the accurate and robust diagnosis, the optimal datum unit has to be decided with much care in featuring anomaly states for robust fault diagnosis of the journal bearing system. The proposed optimal datum unit has

high performance of separation in feature space as well as outperforms in classification prediction results for multi-class problem.

Contribution 2: Definition of new high effective separability metric – PoS

This thesis proposed new separability measure, named PoS. the proposed metric has favorable properties like normalization and boundedness shows a superior ability to evaluate the class separation than conventional metrics like KLD and FDR. Under the multi-class cases, normalization and boundedness of PoS make it possible to compare the separability between different features in the identical manner even with complicated combinations. This implies that these favorable properties can assure the more robust evaluation of separability for non-separable conditions. Since the non-separable conditions are critical in the complex classification problem, it is believed that the optimal datum unit with PoS may also make steps forward to accurately and robustly diagnosing an actual complex engineered system.

Contribution 3: Omnidirectional regeneration of gap sensor signals

This research proposed omnidirectional regeneration (ODR) of gap sensor signals. A journal bearing system generally installs two gap sensors in the right angle with fixed location to obtain the vibration data of the rotor behavior. These fixed sensor sometimes may indicate different signals though the similar health conditions when the directional anomalies like rubbing and misalignment are occurred in different directions. However, the proposed method can generate vibration signals in an

arbitrary direction without using extra sensors. The signals are generated around the circumference of the rotor to consider vibration characteristics in all the directions. The ODR signals guarantee the high performance of reasoning for the journal bearing system.

Contribution 4: Generation of image from vibration signals

This research suggest the generation of image from vibration signals. As mentioned in this study, ODR signals can be generated in circumferential directions. Each regenerated signals are meaningful to the aspect of data for diagnosis. In order to generate vibration image consistently, phase synchronization was applied before ODR. Regenerated signals are stacked in the same order and procedures for construction of two-dimensional image which used as high-dimensional input data for reasoning. Additionally, the stacking of all regenerated vibration signals show the entire behavior of rotor in a glance.

Contribution 5: Extension of diagnosis rule to the actual power plant system

This study propose unsupervised technique for diagnosis of rotor system in a power plant. The diagnosis methodology based on deep learning which has high performance in feature abstraction was implemented by stepwise approach from test-bed to power plant rotor system. These diagnostic framework

7.2 Suggestion of Future Research

Although the technical advances presented in this thesis successfully address critical challenges in diagnostic rules for journal bearing system as well as make the extension of reasoning techniques to the actual power plant, there are still several research areas where further investigations and developments are required to truly bring diagnosis module into reality. Specific suggestion for the continuation of the study on diagnosis of rotor systems are listed as follows:

Optimization of ODR application

This study develops ODR of gap signals to generate virtual vibration data for circumferential directions. However, the optimal number of ODR signal was not defined. Small number of ODR signals is less likely to obtain high accuracy of prediction, whereas too many ODR signals may cause extra computational load without any or little improvement. Therefore, it is required to define appropriate number of ODR which guarantee high performance in classification with the consideration of computational efficiency.

Establishing inverse diagnosis model

The physics of rotor in a journal bearing system were double checked using the FEA model corresponding to the test-bed. FEA model has some advantages in the aspect of time and cost saving compared to the experiment. The development and

improvement the model can increase the fluency of diagnosis methodology. Using the FEA for the anomaly cases which are verified with the experiment, the information of several degree of fault can be obtained. In addition, the information of system condition can be recognized from sensory signals through FEA model inversely, which named inverse diagnosis model.

Advances in diagnostic framework for a power plant using big data

The diagnosis system of a power plant may be faced with a big data issue in the situation of advances in equipment and increasing significance of information. In this regard, deep learning is considered as a good approach for diagnosis of a power plant. As further studies for ease of engineering judgement, there is a need for the integrated diagnostic module between supervised and unsupervised learning which learned from this research. In addition, issues related to big data such as high-dimensional data, computing ability/efficiency and data sampling criteria are also considered as upcoming challenges.

References

[1] L. M. R. Baccarini, V. V. R. E. Silva, B. R. de Menezes, and W. M. Caminhas, "SVM practical industrial application for mechanical faults diagnostic," *Expert Systems with Applications,* vol. 38, pp. 6980-6984, Jun 2011.

[2] L.-l. Jiang, H.-k. Yin, X.-j. Li, and S.-w. Tang, "Fault diagnosis of rotating machinery based on multisensor information fusion using SVM and time-domain features," *Shock and Vibration,* vol. 2014, 2014.

[3] C. Hu, B. D. Youn, T. Kim, and P. F. Wang, "A co-training-based approach for prediction of remaining useful life utilizing both failure and suspension data," *Mechanical Systems and Signal Processing,* vol. 62, pp. 75-90, Oct 2015.

[4] B. D. Youn, K. M. Park, C. Hu, J. T. Yoon, and H. S. Kim, "Statistical health reasoning of water-cooled power generator stator windings against moisture absorption," *IEEE Transactions on Energy Conversion,* vol. PP, pp. 1-10, 2015.

[5] P. Wang, B. D. Youn, C. Hu, J. M. Ha, and B. Jeon, "A probabilistic detectability-based sensor network design method for system health monitoring and prognostics," *Journal of Intelligent Material Systems and Structures,* p. 1045389X14541496, 2014.

[6] P. Wang, B. D. Youn, and C. Hu, "A generic probabilistic framework for structural health prognostics and uncertainty management," *Mechanical Systems and Signal Processing,* vol. 28, pp. 622-637, Apr 2012.

[7] C. Hu, B. D. Youn, P. Wang, and J. T. Yoon, "Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life," *Reliability Engineering & System Safety,* vol. 103, pp. 120-135, 2012.

[8] C. Hu, P. Wang, B. D. Youn, W.-R. Lee, and J. T. Yoon, "Copula-based statistical health grade system against mechanical faults of power transformers," *Power Delivery, IEEE Transactions on,* vol. 27, pp. 1809-1819, 2012.

[9] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical systems and signal processing,* vol. 20, pp. 1483-1510, Oct 2006.

[10] K. Gupta, "Vibration—A tool for machine diagnostics and condition monitoring," *Sadhana,* vol. 22, pp. 393-410, 1997.

[11] J. Lin and L. S. Qu, "Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis," *Journal of Sound and Vibration,* vol. 234, pp. 135-148, Jun 29 2000.

[12] B.-S. Yang and A. Widodo, *Introduction of intelligent machine fault diagnosis and prognosis*: Nova Science Publishers, 2009.

[13] M. P. Norton and D. G. Karczub, *Fundamentals of noise and vibration analysis for engineers*: Cambridge university press, 2003.

[14] K. M. Bossley, R. J. Mckendrick, C. J. Harris, and C. Mercer, "Hybrid computed order tracking," *Mechanical Systems and Signal Processing,* vol. 13, pp. 627-641, Jul 1999.

[15] F. Bonnardot, M. El Badaoui, R. B. Randall, J. Daniere, and F. Guillet, "Use of the acceleration signal of a gearbox in order to perform angular resampling (with limited speed fluctuation)," *Mechanical Systems and Signal Processing,* vol. 19, pp. 766-785, Jul 2005.

[16] L. F. Villa, A. Reñones, J. R. Perán, and L. J. de Miguel, "Angular resampling for vibration analysis in wind turbines under non-linear speed fluctuation," *Mechanical Systems and Signal Processing,* vol. 25, pp. 2157-2168, Aug 2011.

[17] S. J. Loutridis, "Gear failure prediction using multiscale local statistics," *Engineering Structures,* vol. 30, pp. 1214-1223, May 2008.

[18] F. K. Choy, D. H. Mugler, and J. Zhou, "Damage identification of a gear transmission using vibration signatures," *Journal of Mechanical Design,* vol. 125, pp. 394-403, Jun 2003.

[19] H. Yoon and B. D. Youn, "Stochastic quantification of the electric power generated by a piezoelectric energy harvester using a time–frequency analysis under non-stationary random vibrations," *Smart Materials and Structures,* vol. 23, p. 045035, 2014.

[20] Q. He, R. Yan, F. Kong, and R. Du, "Machine condition monitoring using principal

component representations," *Mechanical Systems and Signal Processing,* vol. 23, pp. 446-466, 2// 2009.

[21] Z. Xi, B. D. Youn, and C. Hu, "Random Field Characterization Considering Statistical Dependence for Probability Analysis and Design," *Journal of Mechanical Design,* vol. 132, p. 101008, Oct 2010.

[22] H. Bendjama, S. Bouhouche, and M. S. Boucherit, "Application of Wavelet Transform for Fault Diagnosis in Rotating Machinery," 2012.

[23] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing,* vol. 35, pp. 108-126, Feb 2013.

[24] X. Zhao, T. H. Patel, and M. J. Zuo, "Multivariate EMD and full spectrum based condition monitoring for rotating machinery," *Mechanical Systems and Signal Processing,* vol. 27, pp. 712-728, Feb 2012.

[25] D. Wang, Q. Miao, and R. Kang, "Robust health evaluation of gearbox subject to tooth failure with wavelet decomposition," *Journal of Sound and Vibration,* vol. 324, pp. 1141-1157, Jul 24 2009.

[26] J. Sanz, R. Perera, and C. Huerta, "Fault diagnosis of rotating machinery based on auto-associative neural networks and wavelet transforms," *Journal of Sound and Vibration,* vol. 302, pp. 981-999, May 22 2007.

[27] B. Liu, "Selection of wavelet packet basis for rotating machinery fault diagnosis," *Journal of Sound and Vibration,* vol. 284, pp. 567-582, Jun 21 2005.

[28] B. Jeon, J. Jung, B. D. Youn, Y. Kim, and Y.-C. Bae, "Statistical approach to diagnostic rules for various malfunctions of journal bearing system using Fisher discriminant analysis," in *European Conference of the Prognostics and Health Management Society*, Nante, France, 2014.

[29] T. Han, B.-S. Yang, W.-H. Choi, and J.-S. Kim, "Fault diagnosis system of induction motors based on neural network and genetic algorithm using stator current signals," *International Journal of Rotating Machinery,* vol. 2006, 2006.

[30] B.-S. Yang and K. J. Kim, "Application of Dempster–Shafer theory in fault diagnosis of induction motors using vibration and current signals," *Mechanical*

Systems and Signal Processing, vol. 20, pp. 403-420, 2006.

[31] J. Madhavi and G. V. Yadav, "An Improved Fast Clustering method for Feature Subset Selection on High-Dimensional Data clustering," *International Journal of Application or Innovation in Engineering & Management* vol. 3, 2014.

[32] S. Theodoridis and K. Koutroumbas, *Pattern recognition*, 4th ed.: Elsevier Science & Technology, 2009.

[33] C.-L. Huang and C.-J. Wang, "A GA-based feature selection and parameters optimizationfor support vector machines," *Expert Systems with applications,* vol. 31, pp. 231-240, 2006.

[34] W. Junwen and Z. Xuegong, "A PCA classifier and its application in vehicle detection," in *Neural Networks, 2001. Proceedings. IJCNN '01. International Joint Conference on*, 2001, pp. 600-604 vol.1.

[35] Z. Wanfeng, Y. Yingchun, and Z. Wu, "Exploiting PCA classifiers to speaker recognition," in *Neural Networks, 2003. Proceedings of the International Joint Conference on*, 2003, pp. 820-823 vol.1.

[36] W. Li, T. Shi, G. Liao, and S. Yang, "Feature extraction and classification of gear faults using principal component analysis," *Journal of Quality in Maintenance Engineering,* vol. 9, pp. 132-143, 2003.

[37] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*: John Wiley & Sons, 2012.

[38] Q. P. He, S. J. Qin, and J. Wang, "A new fault diagnosis method using fault directions in Fisher discriminant analysis," *AIChE journal,* vol. 51, pp. 555-571, 2005.

[39] C. S. Tyagi, "A comparative study of SVM classifiers and artificial neural networks application for rolling element bearing fault diagnosis using wavelet transform preprocessing," *Neuron,* vol. 1, pp. 309-317, 2008.

[40] I. A. Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," *Journal of Microbiological Methods,* vol. 43, pp. 3-31, 12/1/ 2000.

[41] J. V. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *Journal of Clinical Epidemiology,* vol. 49, pp. 1225-1231, 11// 1996.

[42] M. Arif, K. A. Alam, and M. Hussain, "Application of Data Mining Using Artificial Neural Network: Survey," *International Journal of Database Theory and Application,* vol. 8, pp. 245-270, 2015.

[43] J. Cannady, "Artificial neural networks for misuse detection," in *National information systems security conference*, 1998, pp. 368-81.

[44] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mechanical Systems and Signal Processing,* vol. 21, pp. 2560-2574, 2007.

[45] J. Shawe-Taylor and N. Cristianini, *Kernel methods for pattern analysis*: Cambridge university press, 2004.

[46] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural computation,* vol. 12, pp. 2385-2404, 2000.

[47] W. L. A. Neves, N. S. D. Brito, B. A. Souza, A. V. Fontes, K. M. C. Dantas, A. B. Fernandes*, et al.*, "Sampling rate of digital fault recorders influence on fault diagnosis," in *Transmission and Distribution Conference and Exposition: Latin America, 2004 IEEE/PES*, 2004, pp. 406-411.

[48] Z. Ping and S. X. Ding, "Influence of Sampling Period on a Class of Optimal Fault-Detection Performance," *IEEE Transactions on Automatic Control,* vol. 54, pp. 1396-1402, 2009.

[49] Y. Lei, Z. He, and Y. Zi, "Application of the EEMD method to rotor fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing,* vol. 23, pp. 1327-1338, 2009.

[50] T. Gałka, "Vibration-Based Diagnostics of Steam Turbines," ed, 2012.

[51] R. C. Eisenmann and R. C. Eisenmann Jr, *Machinery malfunction dignosis and correction*: Prentice Hall PTR, 1997.

[52] D. E. Bently and C. T. Hatch, *Fundamentals of rotating machinery diagnostics*: Amer Society of Mechanical, 2002.

[53] S. N. Ganeriwala, B. Schwarz, and M. H. Richardson, "Operating deflection shapes detect unbalance in rotating equipment," *Sound and Vibration,* vol. 43, pp. 11-13, 2009.

[54] G. Sudhakar and A. Sekhar, "Identification of unbalance in a rotor bearing system," *Journal of Sound and Vibration,* vol. 330, pp. 2299-2313, 2011.

[55] T. Yang and M.-W. Hsu, "An efficient diagnosis technique for variations of shaft-bow and unbalance," in *ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 2009, pp. 57-66.

[56] R. Walker, R. Vayanat, S. Perinpanayagam, and I. Jennions, "Unbalance localization through machine nonlinearities using an artificial neural network approach," *Mechanism and Machine Theory,* vol. 75, pp. 54-66, 2014.

[57] S. Sarkar, A. Nandi, S. Neogy, J. Dutt, and T. Kundra, "Finite element analysis of misaligned rotors on oil-film bearings," *Sadhana,* vol. 35, pp. 45-61, 2010.

[58] T. Gaáka, "Higher harmonic components in steam turbine vibration velocity spectra: a case study," in *Proceedings of the COMADEM*, Faro, Portugal, 2007, pp. 351-359.

[59] I. Redmond, "Shaft Misalignment and Vibration: A Model," *Saudi Aramco J. Technol,* vol. 4, pp. 41-51, 2007.

[60] T. H. Patel and A. K. Darpe, "Vibration response of misaligned rotors," *Journal of Sound and Vibration,* vol. 325, pp. 609-628, 2009.

[61] A. Sekhar and B. Prabhu, "Effects of coupling misalignment on vibrations of rotating machinery," *Journal of Sound and vibration,* vol. 185, pp. 655-671, 1995.

[62] Y.-S. Lee and C.-W. Lee, "Modelling and vibration analysis of misaligned rotor-ball bearing systems," *Journal of Sound and Vibration,* vol. 224, pp. 17-32, 1999.

[63] M. A. Hili, T. Fakhfakh, L. Hammami, and M. Haddar, "Shaft misalignment effect on bearings dynamical behavior," *The International Journal of Advanced*

Manufacturing Technology, vol. 26, pp. 615-622, 2005.

[64] S. Ganesan and C. Padmanabhan, "Modelling of parametric excitation of a flexible coupling–rotor system due to misalignment," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science,* p. 0954406211411549, 2011.

[65] Z. Xiaozhang, D. Xingjian, and L. Ping, "Experimental study of rotor-boundary impact and rub," *Tsinghua Science and Technology,* vol. 2, pp. 700-702, 1997.

[66] H. C. Piccoli and H. I. Weber, "Experimental observation of chaotic motion in a rotor with rubbing," *Nonlinear dynamics,* vol. 16, pp. 55-70, 1998.

[67] F. Chu and W. Lu, "Determination of the rubbing location in a multi-disk rotor system by means of dynamic stiffness identification," *Journal of sound and vibration,* vol. 248, pp. 235-246, 2001.

[68] Q. Wang and F. Chu, "Experimental determination of the rubbing location by means of acoustic emission and wavelet transform," *Journal of Sound and vibration,* vol. 248, pp. 91-103, 2001.

[69] Z. Feng and X.-Z. Zhang, "Rubbing phenomena in rotor–stator contact," *Chaos, Solitons & Fractals,* vol. 14, pp. 257-267, 2002.

[70] F. Chu and W. Lu, "Experimental observation of nonlinear vibrations in a rub-impact rotor system," *Journal of Sound and Vibration,* vol. 283, pp. 621-643, 2005.

[71] M. Farahmandian and M. H. Sadeghi, "Theoretical and Experimental Study on Vibration of Cracked Shafts Using Order Analysis," *International Journal of Advanced Design and Manufacturing Technology,* vol. 7, pp. 77-87, 2014.

[72] M.-C. Wu and S.-C. Huang, "In-plane vibration and crack detection of a rotating shaft-disk containing a transverse crack," *Journal of vibration and acoustics,* vol. 120, pp. 551-556, 1998.

[73] Y.-G. Luo, Z.-H. Ren, H. Ma, T. Yu, and B.-c. Wen, "Stability of periodic motion on the rotor-bearing system with coupling faults of crack and rub-impact," *Journal of Mechanical Science and Technology,* vol. 21, pp. 860-864, 2007.

[74] P. Varney and I. Green, "Crack detection in a rotor dynamic system by vibration monitoring—Part II: extended analysis and experimental results," *Journal of Engineering for Gas Turbines and Power,* vol. 134, p. 112501, 2012.

[75] H. F. De Castro, K. L. Cavalca, and R. Nordmann, "Whirl and whip instabilities in rotor-bearing system considering a nonlinear force model," *Journal of Sound and Vibration,* vol. 317, pp. 273-293, 2008.

[76] C.-C. Fan, J.-W. Syu, M.-C. Pan, and W.-C. Tsao, "Study of start-up vibration response for oil whirl, oil whip and dry whip," *Mechanical Systems and Signal Processing,* vol. 25, pp. 3102-3115, 2011.

[77] Y. Lei, Z. He, and Y. Zi, "A new approach to intelligent fault diagnosis of rotating machinery," *Expert Systems with Applications,* vol. 35, pp. 1593-1600, 2008.

[78] R. G. Kirk and Z. Guo, "Expert system source identification of excessive vibration," *International Journal of Rotating Machinery,* vol. 9, pp. 63-79, 2003.

[79] B.-S. Yang, D.-S. Lim, and A. C. C. Tan, "VIBEX: an expert system for vibration fault diagnosis of rotating machinery using decision tree and decision table," *Expert Systems with Applications,* vol. 28, pp. 735-742, 2005.

[80] W. Meng, F. Guangzeng, and H. Zhenya, "A neural network based fault fuzzy diagnostic system," *Journal of Electronics (China),* vol. 11, pp. 201-207, 1994.

[81] X. Zhang, J. Zhu, D. Qiu, J. Zho, and G. Cao, "Study on the backpropagation neural network applied for fuzzy diagnosis of 200MW turbogenerator set," *Journal of Vibration, Measurement & Diagnosis,* vol. 14, pp. 23-30, 1994.

[82] S. Zhang, T. Asakura, X. Xu, and B. Xu, "Fault diagnosis system for rotary machines based on fuzzy neural networks," in *Advanced Intelligent Mechatronics, 2003. AIM 2003. Proceedings. 2003 IEEE/ASME International Conference on*, 2003, pp. 199-204.

[83] E. De Moura, C. Souto, A. Silva, and M. Irmao, "Evaluation of principal component analysis and neural network performance for bearing fault diagnosis from vibration signal processed by RS and DF analyses," *Mechanical Systems and Signal Processing,* vol. 25, pp. 1765-1772, 2011.

[84] C.-C. Wang, Y. Kang, P.-C. Shen, Y.-P. Chang, and Y.-L. Chung, "Applications of fault diagnosis in rotating machinery by using time series analysis with neural network," *Expert Systems with Applications,* vol. 37, pp. 1696-1702, 2010.

[85] C. Chen and C. Mo, "A method for intelligent fault diagnosis of rotating machinery," *Digital Signal Processing,* vol. 14, pp. 203-217, 2004.

[86] B. Kerezsi and I. Howard, "Vibration fault detection of large turbogenerators using neural networks," in *Neural Networks, 1995. Proceedings., IEEE International Conference on*, 1995, pp. 121-126.

[87] L. M. R. Baccarini, V. V. R. e Silva, B. R. De Menezes, and W. M. Caminhas, "SVM practical industrial application for mechanical faults diagnostic," *Expert Systems with Applications,* vol. 38, pp. 6980-6984, 2011.

[88] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning,* vol. 20, pp. 273-297, 1995.

[89] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *Neural Networks, IEEE Transactions on,* vol. 13, pp. 415-425, 2002.

[90] K. Bacha, S. Souahlia, and M. Gossa, "Power transformer fault diagnosis based on dissolved gas analysis by support vector machine," *Electric Power Systems Research,* vol. 83, pp. 73-79, 2// 2012.

[91] L. Moulin, M. El-Sharkawi, and R. Marks, "Support vector machines for transient stability analysis of large-scale power systems," *Power Systems, IEEE Transactions on,* vol. 19, pp. 818-825, 2004.

[92] L. Deng and D. Yu, "Deep learning: methods and applications," *Foundations and Trends in Signal Processing,* vol. 7, pp. 197-387, 2014.

[93] Y. Bengio, "Learning deep architectures for AI," *Foundations and trends® in Machine Learning,* vol. 2, pp. 1-127, 2009.

[94] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation,* vol. 18, pp. 1527-1554, 2006.

[95] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep

convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[96] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.

[97] G. B. Huang, H. Lee, and E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 2518-2525.

[98] A.-r. Mohamed, G. E. Dahl, and G. Hinton, "Acoustic modeling using deep belief networks," *Audio, Speech, and Language Processing, IEEE Transactions on,* vol. 20, pp. 14-22, 2012.

[99] G. Hinton, D. Li, Y. Dong, G. E. Dahl, A. Mohamed, N. Jaitly*, et al.*, "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups," *Signal Processing Magazine, IEEE,* vol. 29, pp. 82-97, 2012.

[100] D. C. Ciresan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Deep, big, simple neural nets for handwritten digit recognition," *Neural computation,* vol. 22, pp. 3207-3220, 2010.

[101] T. Wang, D. J. Wu, A. Coates, and A. Y. Ng, "End-to-end text recognition with convolutional neural networks," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, 2012, pp. 3304-3308.

[102] D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 3642-3649.

[103] R. Raina, A. Madhavan, and A. Y. Ng, "Large-scale deep unsupervised learning using graphics processors," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 873-880.

[104] G. Hinton, "A practical guide to training restricted Boltzmann machines," *Momentum,* vol. 9, p. 926, 2010.

[105] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the national academy of sciences,* vol. 79, pp. 2554-2558, 1982.

[106] M. A. Carreira-Perpinan and G. E. Hinton, "On contrastive divergence learning," in *Proceedings of the tenth international workshop on artificial intelligence and statistics*, 2005, pp. 33-40.

[107] M. Welling and G. E. Hinton, "A new learning algorithm for mean field Boltzmann machines," in *Artificial Neural Networks—ICANN 2002*, ed: Springer, 2002, pp. 351-357.

[108] T. Tieleman, "Training restricted Boltzmann machines using approximations to the likelihood gradient," in *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 1064-1071.

[109] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," *Advances in neural information processing systems,* vol. 19, p. 153, 2007.

[110] R. Salakhutdinov, "Learning deep generative models," University of Toronto, 2009.

[111] *Mechanical vibration Evaluation of machine vibration by measurements on rotating shafts,* I. O. f. Standardization ISO 7919-2, 2001.

[112] I. Ansys, "Rotordynamic analysis guide," *Swanson Analysis Systems, Inc, Tech. Rep,* 2009.

[113] K. B. Kwon, J. S. Han, B. C. Jeon, J. H. Jung, and B. D. Youn, "Rotordynamic analysis and validation of RK4 rotating kit to detect

abnormal conditions," *Proceedings of the KSNVE Annual Autumn Conference,* pp. 994-997, 2014.

[114] Y. Lei, Z. He, and Y. Zi, "Application of an intelligent classification method to mechanical fault diagnosis," *Expert Systems with Applications,* vol. 36, pp. 9941-9948, Aug 2009.

[115] J. Lin and M. J. Zuo, "Gearbox fault diagnosis using adaptive wavelet filter," *Mechanical Systems and Signal Processing,* vol. 17, pp. 1259-1269, Nov 2003.

[116] B. Sreejith, A. Verma, and A. Srividya, "Fault diagnosis of rolling element bearing using time-domain features and neural networks," in *Industrial and Information*

Systems, 2008. ICIIS 2008. IEEE Region 10 and the Third international Conference on, 2008, pp. 1-6.

[117] H. C. Sun, C. M. Huang, and Y. C. Huang, "Fault Diagnosis of Steam Turbine-Generator Sets Using an EPSO-Based Support Vector Classifier," *Ieee Transactions on Energy Conversion,* vol. 28, pp. 164-171, Mar 2013.

[118] O. Ludwig, U. Nunes, R. Araújo, L. Schnitman, and H. A. Lepikson, "Applications of information theory, genetic algorithms, and neural models to predict oil flow," *Communications in Nonlinear Science and Numerical Simulation,* vol. 14, pp. 2870-2885, Jul 2009.

[119] O. Ludwig and U. Nunes, "Novel maximum-margin training algorithms for supervised neural networks," *Neural Networks, IEEE Transactions on,* vol. 21, pp. 972-984, Jun 2010.

[120] L. J. Eshelman and J. D. Schaffer, "Real-coded genetic algorithms and interval-schemata," 1992.

[121] C. C. Chang and C. J. Lin, "LIBSVM: A Library for Support Vector Machines," *Acm Transactions on Intelligent Systems and Technology,* vol. 2, p. 27, 2011.

[122] S. Kullback and R. A. Leibler, "On information and sufficiency," *The Annals of Mathematical Statistics,* pp. 79-86, 1951.

[123] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*: Elsevier Science, 2008.

[124] M. L. Roush and W. M. Webb, *Applied reliability engineering* vol. 2: Center for Reliability Engineering, University of Maryland, 2006.

[125] B. F. Guo, R. I. Damper, S. R. Gunn, and J. D. B. Nelson, "A fast separability-based feature-selection method for high-dimensional remotely sensed image classification," *Pattern Recognition,* vol. 41, pp. 1653-1662, May 2008.

[126] M. Afgani, S. Sinanovic, and H. Haas, "Anomaly detection using the Kullback-Leibler divergence metric," in *Applied Sciences on Biomedical and Communication Technologies, 2008. ISABEL'08. First International Symposium on*, 2008, pp. 1-5.

[127] D. H. Johnson, C. M. Gruner, K. Baggerly, and C. Seshagiri, "Information-theoretic

analysis of neural coding," *Journal of computational neuroscience,* vol. 10, pp. 47-69, 2001.

[128] S. S. Rao, *Reliability-based design*: McGraw-Hill Companies, 1992.

[129] M. I. Friswell, *Dynamics of rotating machines*: Cambridge University Press, 2010.

[130] K. N. Gupta, "Vibration — A tool for machine diagnostics and condition monitoring," *Sadhana,* vol. 22, pp. 393-410, 1997/06/01 1997.

[131] Y. Wu, S. Li, S. Liu, H.-S. Dou, and Z. Qian, "Vibration-based condition monitoring," in *Vibration of Hydraulic Machinery*, ed: Springer, 2013, pp. 431-477.

[132] M. W. Gardner and S. R. Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," *Atmospheric Environment,* vol. 32, pp. 2627-2636, 8/1/ 1998.

[133] A. K. Jain, M. Jianchang, and K. M. Mohiuddin, "Artificial neural networks: a tutorial," *Computer,* vol. 29, pp. 31-44, 1996.

[134] K. F. Thang, R. K. Aggarwal, A. J. McGrail, and D. G. Esp, "Analysis of power transformer dissolved gas data using the self-organizing map," *Power Delivery, IEEE Transactions on,* vol. 18, pp. 1241-1248, 2003.

[135] M. C. Kind and R. J. Brunner, "SOMz: photometric redshift PDFs with self-organizing maps and random atlas," *Monthly Notices of the Royal Astronomical Society,* vol. 438, pp. 3409-3421, 2014.

Abstract (Korean)

대형 발전소 저널베어링 회전체 시스템의 비지도 학습 및 진단 기법 연구

전 병 철
서울대학교 대학원
기계항공공학부

회전체 시스템은 다양한 산업용 기계 및 설비에 널리 사용되고 있으며, 종종 시스템의 고장으로 인해 상당한 경제적 손실이 초래된다. 특히, 저널베어링이 일반적으로 사용되는 발전소와 같은 산업분야에서는 설비의 고장이 상당히 중요한 문제로 다루어진다. 따라서 회전체 시스템에 대한 진단 도구들이 일반적으로 구축된다. 데이터에 기반한 자동화 진단 시스템은 초기 단계의 비정상을 감지함으로써 시스템의 안전성 증대와 함께 비용절감에 기여할 수 있다. 진단 알고리즘을 개발하여 실제 시스템에 적용하기 위해서는 진단 기법의 강건성 확보가 무엇보다 중요하며, 나아가 실제 환경에서 발생할 수 있는 다양한 고장 조건을 진단하기 위한 비지도 학습에 대한 고려가 필요하다.

본 연구에서는 저널베어링 회전체 시스템의 강건한 진단 기법 개발을 위한 두 가지 중요한 연구 내용으로 진단용 데이텀 기준 최적화와 변위 센서의 전방향 신호 재생성을 다루고 있다. 데이텀 기준에 대한 연구에서는 특성인자 공간에서의 분류 능력과 이상상태 진단 결과에 대한 평가를 통해 최적의 조건이 정의된다. 전방향 신호 생성 연구에서는 회전체의 전체적인 거동 특성을 충분히 반영할 수 있는 신호 생성 기법을 제시함으로써 높은 진단 정확성과 강건성 확보가 가능하다. 이상의 연구에서 개발된 강건한 진단 기법을 실제 시스템에 적용하기 위해 비지도 학습을 이용한 진단 기법이 고려된다. 딥 러닝을 이용한 비지도 학습은 고차원 특성인자 추출을 통해 우수한 진단 성능을 나타내며, 발전소 진단 프레임워크로 적용 가능성을 제시한다.

주제어 : 진단

데이텀 기준

전방향 신호 재생성

저널 베어링

딥 러닝

학 번 : 2012-30730