Master's Thesis

An abrupt variance analysis of multiple sensor signals for dimension reduction in fault diagnosis and prognosis

Ha Young Oh

Department of System Design and Control Engineering

Graduate School of UNIST

2018

An abrupt variance analysis of multiple sensor signals for dimension reduction in fault diagnosis and prognosis

Ha Young Oh

Department of System Design and Control Engineering

Graduate School of UNIST

An abrupt variance analysis of multiple sensor signals for dimension reduction in fault diagnosis and prognosis

A thesis
submitted to the Graduate School of UNIST
in partial fulfillment of the
requirements for the degree of
Master of Science

Ha Young Oh

12. 19. 2017
Approved by

Advisor
Duck Young Kim

ULSAN NATIONAL INSTITUTE OF SCIENCE AND TECHNOLOGY

An abrupt variance analysis of multiple sensor signals for dimension reduction in fault diagnosis and prognosis

Ha Young Oh

This certifies that the thesis of Ha Young Oh is approved.

12. 19. 2017

Advisor: Duck Young Kim

Daeil Kwon

Youngshin Kwak

Abstract

Many studies related to condition based maintenance (CBM) have been conducted especially for quality monitoring in motor, shipbuilding and electronics industries and equipment diagnosis in large-scale plant or automation machines.

Sensor data related to critical components are collected using many sensors to analyze complex system or equipment. When conducting fault diagnosis using high dimensional time series data composed of many sensors, pre-processing steps such as selecting sensors related to system failure is needed for effective analysis. In other words, selecting sensors is a kind of process to reduce dimension of multivariate data.

Many researchers have studies dimension reduction techniques for hundreds of years. Among many dimension reduction techniques, Principal Component Analysis, Linear Discriminant Analysis, and Partial Least Squares are widely used methods. PCA, which is a bible in dimension reduction techniques, basically uses variation of each sensor to decide new principal components, which is newly made axes. However, due to these intrinsic characteristic emphasizing variance, sensor of which signal is highly fluctuating periodically can be ranked as a highly important sensor even though it does not have any relation with system failure. That is, there is a limit to improve fault diagnosis algorithm directly using PCA sensor selection since it only considers total variance of data not finding principal sensors distinguishing fault and no-fault state.

Therefore, in this study (i) we discuss key characteristics of sensor signals which are effective to distinguish no-fault and fault state of a system and introduce indices considering those characteristics: abrupt variance, discernibility index, and sparse impulse, (ii) propose sensor selection methods considering proposed indices and (iii) propose new principal component using abrupt variance-based PCA. The proposed sensor selection methods is illustrated and demonstrated with the case studies of vehicle fault simulator and gear fault simulator.

Table of Contents

List of Figures

List of Tables

I. Introduction

1.1 Background

Mechanical defects cause a lot of loss such as facility stop and bad product quality. Thus, many studies regarding maintenance have been conducted and progressed to increase the reliability of the system and reduce operational cost (Xiao *et al.*, 2013; Ying *et al.*, 2010).

Thus, operation and maintenance (O&M) techniques have been progressed. Starting from corrective maintenance (Kenne & Boukas, 1997), preventive maintenance (Malik, 1979), condition based monitoring such as statistical process control (SPC) (Kano & Nakagawa, 2008), and predictive maintenance (Lu et al., 2009) have been studied and even needs for predictive process adjustment arose.

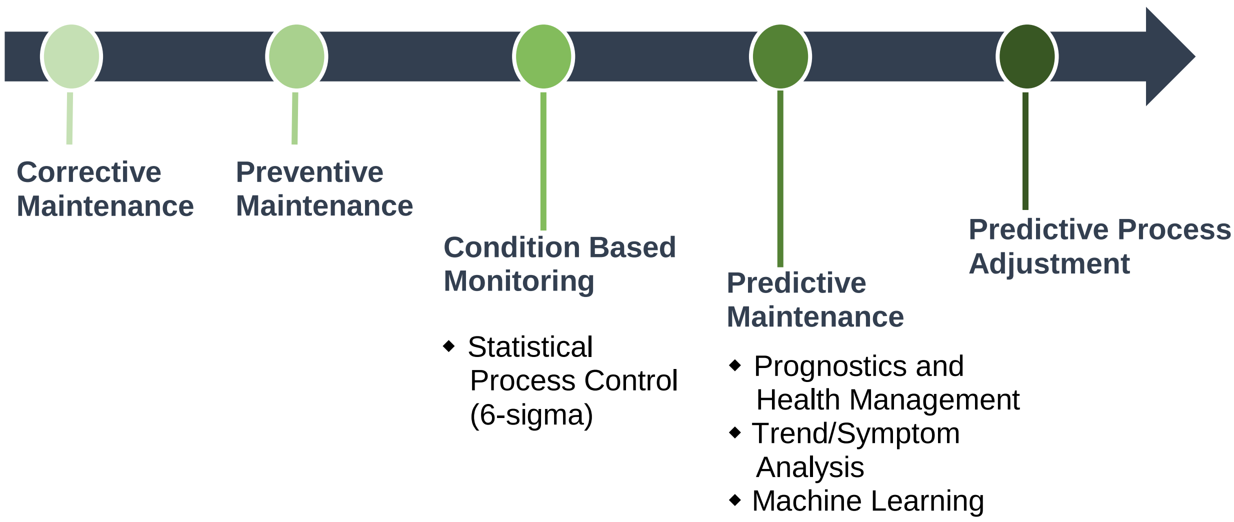


Figure I-1 Progress in operation and maintenance techniques

As for condition based monitoring techniques which have been already widely used in various industrial fields, univariate SPC, multivariate SPC were simply used to fault diagnosis and prognosis and more advance techniques are methods using pattern mining or machine learning algorithm such as support vector machine, neural network and so on. Apart from detail methods, overall goal of fault diagnosis and prognosis researches is improving performance considering accuracy and computational time.

Overall structure of fault diagnosis is shown in the Figure I-2. From the machinery, sensor data which can show machinery conditions are collected using data acquisition unit. Then, signals are pre-processed to increase efficiency of next step. Pre-processing includes noise removal, dimension reduction and so on. Pre-processed data are then used to extracting meaningful features for classification and extracted features are used in fault diagnosis at last.

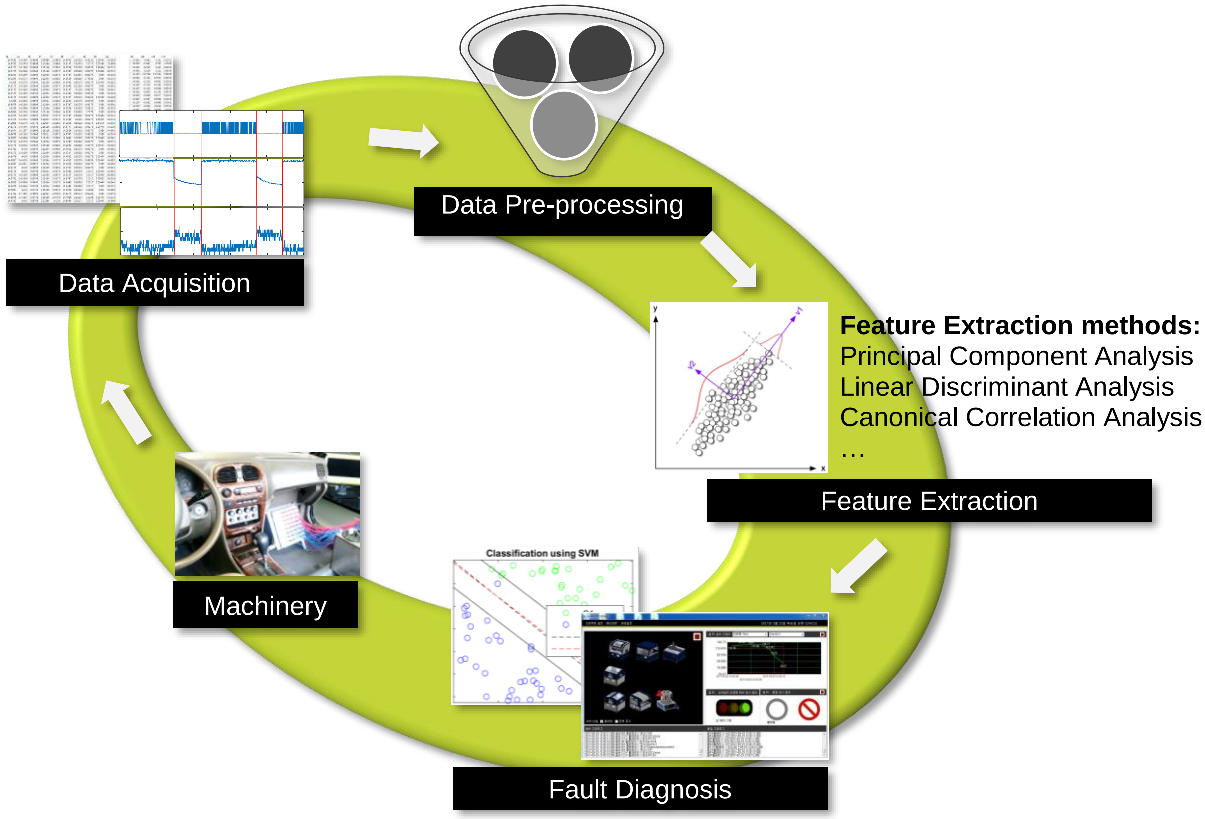


Figure I-2 Overall structure of fault diagnosis

Distinguishing fault and no-fault state is classification problem. Thus, many studies have done to improve classification performance to improve overall diagnostic performance. For example, SagHa proposed a method to detect anomalous sensors in sensor networks by recognizing relationship between each sensors and networks. Then, anomalous sensors are removed in classifier fusion process. The result shows classification accuracy was improved by this method (Sagha *et al.*, 2011). Also, wavelet and SVM were used to detect bearing fault of induction motor (Konar & Chattopadhyay, 2011). In summary, studies to improve classification result have been done in almost all fault diagnosis stages such as pre-processing stage and feature extraction stage.

1.2 Motivation

As mentioned earlier, there have been many methods and techniques studied to develop classification performance. Among them, dimension reduction is one of the powerful and conventional methods of data pre-processing by selecting sensors among all dataset. Dimension reduction can be divided into two sections such as sensor selection conducted in transformed space, which means new space in reduced domain and sensor selection by variable subset selection.

Among them, in this paper for clear understanding and interpretation of data, data which is transformed in new space is not preferred. For instance, considering vessel diesel engine there exists more than fifty sensors. In this case, several dimension reduction techniques such as Principal Component Analysis (PCA) can give new reduced axes which named principal components. Basic principle of PCA is based on the variance, finding the axis which can maximize total variance of data. However, in practical environment, it is hard to interpret meaning of derived axis and even if the axis is named, newly made axis might be not meaningful by states of the system.

Another method exists when selecting important sensors among dimension reduction techniques. With sensor selection methods using variable subset selection, original sensor name and information is not transformed into new space. Instead, original sensor subsets are selected. However, when conducting sensor selection by variable subset selection simple classification algorithms are used so that it takes a time to get selected sensor list.

Thus, new sensor selection method is necessary for getting computational efficiency and increasing detection accuracy simultaneously.

1.3 Objectives

To enhance the performance of fault diagnosis and prognosis, machinery monitoring systems usually use rich information from multiple sensors. For example, in vehicle engine fault generator there exists forty sensors to monitoring condition of engine such as injector, O2 sensor, fuel pump relay control and so on. However, monitoring of machine, it takes time to process data with too many sensors, even though rich information can have much more potentially related interpretation for classifying fault and no-fault state of a system. Also, it is not always the case that classification using all original data, which is m dimensional, results better performance. Sometimes, total data with all sensors might have redundant data.

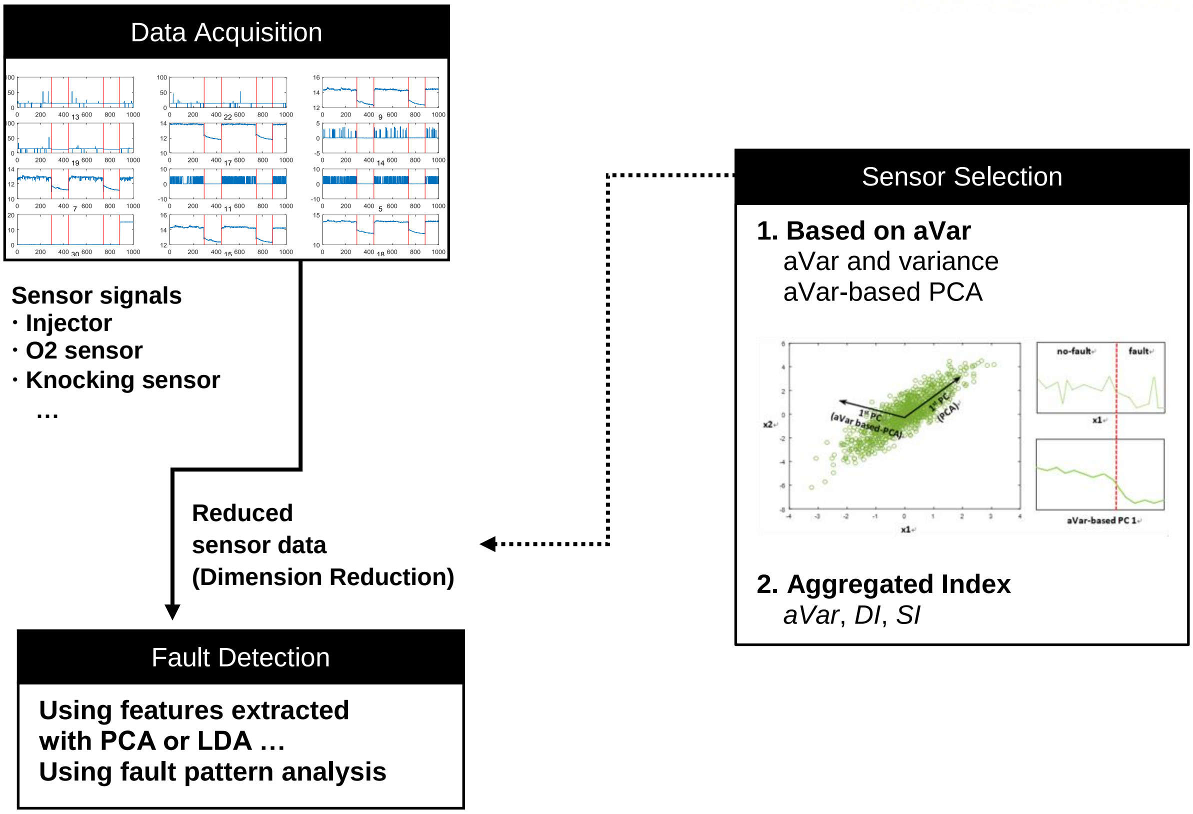


Figure I-3 A framework for fault detection using sensor selection methods

In this regard, this research aims to first, discuss key characteristics of sensor signals which are effective to classify machine states and second, to propose sensor selection methods based on the signal characteristics discussed. Figure I-3 illustrates the overall framework for a fault detection using sensor selection techniques, and which consists of two main stages: (i) the online monitoring system for machinery fault diagnosis and (ii) sensor selection stage, which is done in offline training stage.

1.4 Outline of the thesis

This thesis consists of five Chapters. Chapter 1 introduces brief background, motivation and objectives of this paper. In the Chapter 2, literature reviews, which consist of dimension reduction techniques especially related to sensor selection and characteristics of sensor signals, are presented. The proposed key characteristics of sensor signals and way to select sensors are described in Chapter 3 and case studies which can prove efficiency of the proposed method are presented in Chapter 4. Lastly, the conclusions and future research are described in Chapter 5.

II.Literature survey

2.1 Dimension reduction techniques

Since dimension reduction of high dimensional data is highly required in industrial fields, many researchers have studied dimension reduction techniques for more than hundreds of years.

Dimension reduction methods can be mainly divided into two as feature selection and feature extraction by several researchers. In feature extraction, original feature space is projected onto new feature space by reducing high dimensional original data. New feature space can be represented as a linear combination of original features. One of famous feature extraction methods is Principle Component Analysis (Wold *et al.*, 1987) which has been widely and mainly used in dimension reduction studies. Furthermore, Linear Discriminant Analysis (Izenman, 2013), Partial Least Squares (Barker & Rayens, 2003), Canonical Correlation Analysis (Thompson, 2005) are also widely used methods in feature extraction. In second method named feature selection methods, subset of original features which have strong relationship to the model are selected so that total data are explainable using selected subset at most. Well known feature selection methods are Information Gain (Xing *et al.*, 2001), Relief (Kira & Rendell, 1992), Laplacian Score (He *et al.*, 2006), Fisher Score (Gu *et al.*, 2012), Lasso methods (Tang *et al.*, 2014) and so on.

Table II-1 Variable selection methods and dimension reduction methods (Hartmann, 2004)

Variable selection methods	Dimension reduction methods
Exploratory modeling (unsupervised learning) · coloring a correlation matrix · sparse principal components · Principal variables · Methods of discrete variable clustering	Exploratory modeling (unsupervised learning) · (kernel) PCA and factor analysis · singular value decomposition · and correspondence analysis · multidimensional scaling · methods of fuzzy variable clustering
Predictive modeling (supervised learning) · type 3 analyses · subset selection in regression (stepwise regression) · recursive partitioning and regression trees · step-up and step-down multivariate testing	Predictive modeling (supervised learning) · (kernel) partial least squares · sliced inverse regression and principal hessian direction · neural networks and support vector machines

Variable selection methods	Dimension reduction methods
· Garotte by Breiman · (univariate) soft thresholding · Lasso · Elastic net · sparse L1 SV regression · sparse L1 SV classification · SVM feature selection by Guyon · Feature selection using genetic algorithm · Bayesian methods of variable selection	

However, terms for dividing dimension reduction methods vary according to even professional researchers. Hartmann even tried to classify variable selection methods and dimension reduction methods clearly in order to emphasize the difference between two methods. Variable selection methods and dimension reduction method could be summarized in the Table II-1.

Also, new approach for variable selection with dimensionality reduction was proposed by Lior Wolf (Wolf & Bileschi, 2005). The author emphasized the importance of feature selection by comparing conventional PCA approach and the proposed method. In his method, variable selection algorithm is applied after informative features from the data are extracted. Then, dimension reduction algorithms are applied to extract the vector which can represent the data well by keeping the same optimization function on the dimension reduction and feature selection stage.

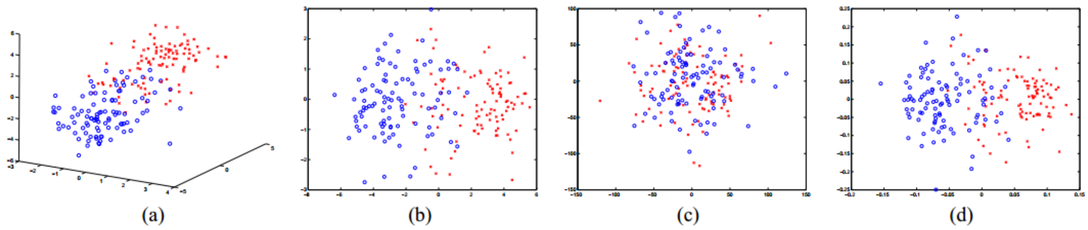


Figure II-1 The importance of feature selection. Each algorithm differs by the way dealing with irrelevant variables (a): choosing three relevant variables among 203 dimensions (b): The first 2 PCs of three relevant dimensions (c): PCA result of the whole dimensions (d): The result of applying PCA with weights (Wolf & Bileschi, 2005)

Table II-2 Clarified terminologies used in this paper

Terms used in different forms	Terms used in this paper
· Feature selection · Sensor selection · Variable selection	· Sensor selection by variable subset selection
· Variable selection in feature extraction · Sensor selection in feature extraction	· Sensor selection by space transformation

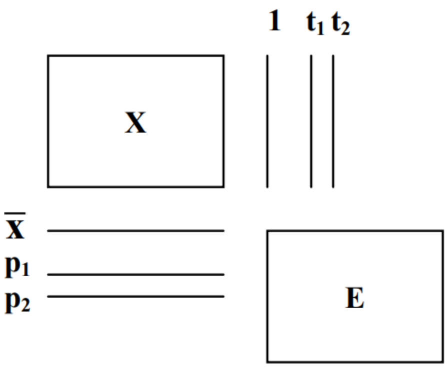
Feature selection is also known as variable selection, attribute selection or variable subset selection and so on. To clarify the terms used in the paper, we will use ‘sensor selection by variable subset selection’ instead of variable selection and feature selection, also use ‘sensor selection by space transformation’ instead of feature extraction for preventing the confusion that might occur due to different usage of term by people to people like in the Table II-2.

2.1.1 Sensor selection by space transformation

As already mentioned above, sensor selection methods are divided into two subsections. Among them, first section is a sensor selection which can be done in transformed space. Principal Component Analysis, Linear Discriminant Analysis, Partial Least Squares are widely and traditionally used techniques among them. As a result of each method, new axis which could explain origial dataset in lower dimensions are generated. The effectiveness and usefulenss of these methods have been fully verified already but they have weakness in terms of interpretation and felxibility in usage such as change in state. Among them, two main methods, Principal Component Analysis and Linear Discriminant Anlaysis are discussed further in this section. Basic principle of each methods and way to select sensors are handled.

Principal Component Analysis (PCA)

Principal Component Analysis is well known and widely used methods in dimension reduction. The method minimizes total sum of error between projected data and original data by capturing maximum variance direction in data matrix X.

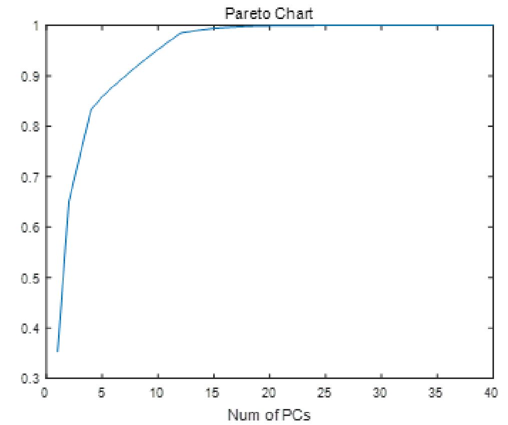


$$X = 1 * \bar{X} + TP + E$$

Figure II-2 A principle of PCA. X-data matrix, T-score matrix, P-loading matrix, E-residual matrix, $\bar{X}$-mean value of the columns, 1-a vector of ones. (Mörtsell & Gulliksson, 2001)

Conventional PCA results principal components which are newly generated axis in reduced dimension. However, when using principal components not original sensors it might be good for dimension reduction but it is hard to clarify and interpret the real meaning of them.

There are several preliminary options when conducting dimension reduction especially in the case of PCA. First, user should determine the number of factors to retain. In practice, user decide the number arbitrarily or based on the threshold (e.g. threshold>90%, threshold>80%). Without considering relevance or redundancy or data, it is always good to use many sensors as possible. Rich data can contain all information which might be even potentially effective to fault detection even though it does not looks like effective in mono criteria. Thus, it is upto users deciding the number of sensors used having assupmtion that the more the sensors exist, the better the classifcation performance would be.



	PC 1	PC 2	PC 3		PC K			PC N	SUM(1:K)
var 1	0.0062	0.0047	0.0062		-0.0020	0.0006	0.1735	0.0739	0.2632
var 2	0.0058	-0.0007	0.0098		-0.0083	-0.0136	0.6930	-0.6717	0.0144
var 3	-0.0013	-0.0001	0.0004		0.0007	0.0009	-0.0009	0.0006	0.0004
var 4	-0.0035	-0.0002	0.0041		-0.0038	0.0017	-0.0146	0.0035	-0.0128
var 5	-0.0003	-0.0563	0.8602		-0.5062	-0.0097	-0.0110	0.0082	0.2849
var 6	0.0045	-0.0005	0.0055		0.0008	0.0047	0.0033	-0.0009	0.0175
	-0.0473	0.7048	0.0309		-0.0362	-0.0051	0.0032	0.0019	0.6523
	-0.0015	0.0003	-0.0029		-0.0007	0.0012	0.0218	0.0109	0.0291
	-0.0009	0.0001	0.0007		0.0003	0.0013	0.0442	0.0226	0.0683
	0.0041	0.0013	0.0048		0.0003	-0.0197	0.0738	-0.0405	0.0240
	0.0051	-0.0006	0.0029		-0.0035	-0.0015	0.0034	-0.0011	0.0049
⋮	0.0513	-0.7031	-0.0181	…	0.0368	0.0056	0.0033	0.0006	-0.6236
	0.0066	-0.0018	-0.0014		-0.0011	0.0042	0.4820	0.6951	1.1837
	-0.0024	0.0013	0.0012		0.0009	0.0017	0.4085	0.1922	0.6034
	-0.0002	-0.0002	0.0007		0.0018	0.0004	0.0005	0.0016	0.0047
	-0.0004	-0.0006	0.0016		0.0041	-0.0004	0.0002	0.0006	0.0052
	-0.0001	-0.0005	0.0038		-0.0006	0.0124	0.2861	0.1427	0.4438
	-0.0008	-0.0005	0.0023		-0.0002	0.0011	-0.0006	-0.0006	0.0007
	0.0050	-0.0001	0.0021		-0.0023	-0.0021	-0.0035	-0.0038	-0.0047
	-0.0006	-0.0048	0.0064		0.0001	0.0059	0.0002	-0.0003	0.0069
	-0.0011	0.0015	0.0062		-0.0030	-0.0006	-0.0001	0.0004	0.0032
	0.0020	-0.0004	0.0027		-0.0014	0.0005	0.0017	-0.0013	0.0038
	-0.0009	-0.0002	0.0000		-0.0004	0.0001	-0.0004	-0.0004	-0.0023
var N	-0.0010	-0.0005	-0.0001		-0.0009	0.0002	-0.0003	-0.0001	-0.0027

Figure II-3 Sensor ranking using standard PCA

Second subject is the way to select sensors based on PCA. It differs by practitioners or researchers. Some researchers select sensors only considering coefficient value of first major principal component. Otherwise, some researchers select sensors based on summation of coefficients. For better understanding, sensor ranking methods are shown in the Figure II-3. For the first step, based on the threshold value which were set by users, the number of principal components, which is k in this example, attained are decided. After that, coefficients are summed from first column to k^{th} column. By ordering summed coefficients values, sensors are ranked and top k^{th} sensors are selected finally.

In this paper, basic assumption is that the number of principal component selected or the number of sensors selected are user-defined. However, there have been a variety of opinions in terms of the way to select the number of sensors. Some researchers thought user-defined value is too naïve and uncertain. Then, for these needs , in order to make this uncertain criteria certain, there have been several studies done to try to find out which factors have to be retained. Among them, one study had highlighted there exists optimal threshold. Thus, optimal hard threshold was proposed. After estimating noise of dataset, author removed components which have singular value less than specific threshold defined in this paper (Donoho & Gavish, 2013). In other paper, author used Bayesian model selection to estimate data dimension by interpreting PCA in terms of density estimation (Minka, 2001). In other words, Bayesian model selection method was applied to probabilistic PCA. In order to decide a dimension of a specific subspace, probability of data of each dimension was calculated and the maximum value was selected.

Also, there exists another method to select variable using principal component anlaysis. Yifie Wang summarized three variable selection methods based on the principal components such as B2 method, B4 method and H method (Y. Wang & Ma, 2012). In B2 method, for the first step the principal components with eigenvalue less than a predefined constant are selected. Then starting with eigenvectors with smallest eigenvalue, variable of which eigenvector has the largest eigenvalue is eliminated until all the eigenvectors are examined. Different from B2 method, the selected variables have largest absolute coefficient value in B4 method. Third method is H method which author suggested the best among three methods. Basic principle of way to select variables are similar to B2 and B4 methods. However, In H method new index h, which is the sum of the squared correlations between variables are examined until reaching the threshold.

In summary, there are several methods selecting sensors and ranking sensors using Principal Component Analysis. Each methods has its own characteristics and strengths and weaknesses. Thus, it is necessary to compare those characteristics and select appropricate methods to application.

Linear Discriminant Analysis (LDA)

Second methods which can be used to select sensors in transformed space is Linear Discriminant Analysis. Fundamental objective of PCA and LDA is quite different. Fundamental objective of PCA is find out new principal components. Selecting original sensors is not original purpose of PCA. Fundamental principle of LDA is classification. In other words, Linear Discriminant Analysis generally focuses on the classification only whereas PCA focuses on the signal representation referring the Figure II-4. Also, similar with PCA the main purpose of LDA is not sensor selection. It just provides classification results.

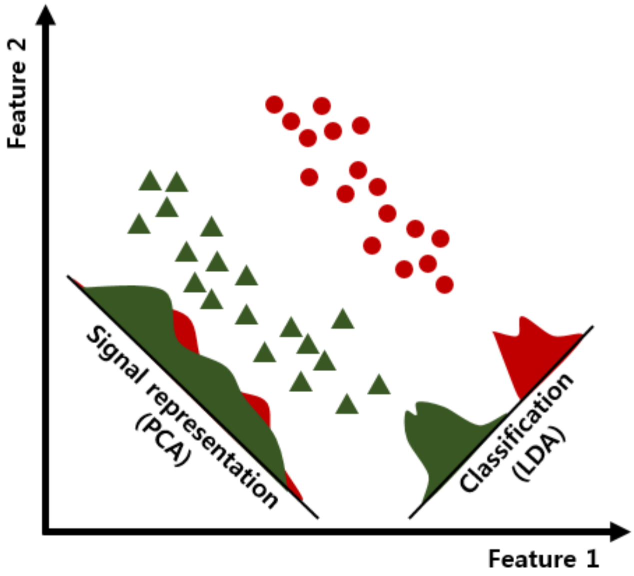


Figure II-4 Comparison between PCA and LDA

Trendafilov and Jolliffe suggested variable selection method using LASSO constraints in discriminant analysis named DALASS (Trendafilov & Jolliffe, 2007). They solved the classical discriminant analysis problem adding LASSO constraints. However, unlike Principal Component Analysis, a basis is not provided by the discriminant function coefficients so it is hard to get some insight from unique and simple interpretation.

Likewise, two main methods of sensor selection in transformed space were discussed. In both cases, the sensor selection, which is selecting reduced number of original variables among total original variables, is not fundamental and main purpose so that the studies which handles the way to select original sensors are relatively few.

2.1.2 Sensor selection by variable subset selection

In variable subset selection methods, best sensors can be found by removing features that are not relevant to the model or that are redundant. The main point of the variable subset selection is not generating new representation of the data. In other words, it keeps original representation of the data (Saeys *et al.*, 2007). By doing so, sensors can be selected without any transformation and the physical meanings of the original dataset are maintained so that it is superior in terms of interpretability than sensor selection using in transformed space.

For clear understanding of variable subset selection, there exists many literature survey papers. Saeys reviewed each method and compared its advantages and disadvantages. Detail categorization and explaination is in the Figure II-5.

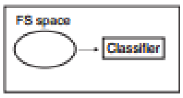
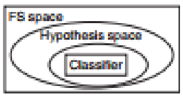
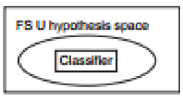
Model search	Advantages	Disadvantages
Filter	Univariate	
	Fast Scalable Independent of the classifier	Ignores feature dependencies Ignores interaction with the classifier
	Multivariate	
	Models feature dependencies Independent of the classifier Better computational complexity than wrapper methods	Slower than univariate techniques Less scalable than univariate techniques Ignores interaction with the classifier
Wrapper	Deterministic	
	Simple Interacts with the classifier Models feature dependencies Less computationally intensive than randomized methods	Risk of over fitting More prone than randomized algorithms to getting stuck in a local optimum (greedy search) Classifier dependent selection
	Randomized	
	Less prone to local optima Interacts with the classifier Models feature dependencies	Computationally intensive Classifier dependent selection Higher risk of overfitting than deterministic algorithms
Embedded 	Interacts with the classifier Better computational complexity than wrapper methods Models feature dependencies	Classifier dependent selection

Figure II-5 A taxonomy of feature selection techniques. For each feature selection type, author highlighted a set of characteristics which can help to choose a technique (Saeys *et al.*, 2007)

Sensor selection method using variable subset selection are mainly divided into three methods such as filter methods, wrapper methods, and embedded methods. More detail explanation for each method will be handled in following sections.

Filter methods

Filter methods can provide a generic selection of variables without using classification algorithm, which is a learning machine. This method evaluates the goodness of selected subsets by observing the intrinsic signal characteristics such as mean, RMS and so on. Thus, filter methods could be regarded as a preprocessing step for dimension reduction.

INPUT:
$D = \{X, L\}$ — *// a training data set with n number of features where*
// $X = \{f_1, f_2, f_3, \ldots, f_n\}$ *and L labels*
X' — *// predefined initial feature subset* ($X' \subset X$ *or* $X' = \{\phi\}$)
θ — *// a stopping criterion*
OUTPUT: X'_{opt} — *// an optimal subset*

Begin:
Initialize:
  $X_{opt} = X'$;
  $\varphi_{opt} = E(X', I_m)$; *// evaluate* X' *by using an independent measure* I_m
do begin
  $X_g = generate(X)$; *// Subset generation for evaluation*
  $\varphi = E(X_g, I_m)$; *//* X_g *current subset evaluation by* I_m
  If $(\varphi > \varphi_{opt})$
    $\varphi_{opt} = \varphi$;
    $X'_{opt} = X_g$;
repeat *(until* θ *is not reached);*
end
return X'_{opt};
end;

Figure II-6 A general explanation for filter algorithm (Kumar & Minz, 2014)

A general filter algorithm could be explained as in the Figure II-6. Before meeting a stopping criterion, subset which were generated are repeatedly used in evaluating the performance of classification. In filter algorithm feature dependencies are not considered so that its main advantage is computational efficiency.

There are some examples in filter algorithms such as correlation-based feature selection, the fast correlation-based filter method, information gain and ReliefF and so on.

Wrappers

Second methods of variable subset selection is wrapper method. Different from the filter algorithms, feature dependencies and interactions are considered in wrapper methods. Figure II-7 shows overall algorithms of wrapper methods.

For instance, in a wrapper method which randomly selected sensor subsets, it takes time to evaluate the classification performance using randomly generated subsets. Compared to filter methods, it is computationally inefficient.

INPUT:
$D = \{X, L\}$ *// a training data set with* n *number of features where*
// $X = \{f_1, f_2, f_3, \dots, f_n\}$ *and L labels*
X' *// predefined initial feature subset* $(X' \subset X or X' = \{\phi\})$
θ *// a stopping criterion*
OUTPUT: X'_{opt} *// an optimal subset*

Begin:
Initialize:
$X_{opt} = X'$;
$\varphi_{opt} = E(X', A)$; *// evaluate* X' *by using mining algorithm A*
do begin
$X_g = generate(X)$; *// Subset generation for evaluation*
$\varphi = E(X_g, A)$; *//* X_g *current subset evaluation by A*
If $(\varphi > \varphi_{opt})$
$\varphi_{opt} = \varphi$;
$X'_{opt} = X_g$;
repeat *(until* θ *is not reached);*
end
return X'_{opt};
end;

Figure II-7 A general explanation for wrapper algorithm (Kumar & Minz, 2014)

Embedded techniques

Third method is embedded technique which considers both advantages and disadvantages of filter methods and wrapper methods. One example of embedded techniques is random forest. It takes comparatively lower computational cost than wrapper methods. Also, it interacts with learning algorithms and considers feature dependencies. The Figure II-8 helps understanding embedded techniques.

INPUT:
$D = \{X, L\}$ *// a training data set with n number of features where*
// $X = \{f_1, f_2, f_3, \ldots, f_n\}$ *and L labels*
X' *// predefined initial feature subset* $(X' \subset X or X' = \{\phi\})$
θ *// a stopping criterion*
OUTPUT: X'_{opt} *// an optimal subset*

Begin:
Initialize:
$X_{opt} = X'$;
$\varphi_{opt} = E(X', I_m)$;   *// evaluate* X' *by using independent evaluation measure*
$\delta_{opt} = E(X', A)$;   *// evaluate* X' *by using mining algorithm A*
$C_0 = C(X')$;   *// cardinality calculation of* X'
do begin
for $k = C_0 + 1$ *to* n
for $i = 0$ *to* $n - k$
$X_g = X_{opt} \cup \{f_i\}$;   *// Subset generation for evaluation with cardinalityk*
$\varphi = E(X_g, I_m)$;   *// evaluation the current subset* X_g *by* I_m
If $(\varphi > \varphi_{opt})$
$\varphi_{opt} = \varphi$;
$X'_{opt} = X_g$;
end
$\delta = E(X'_{opt}, A)$;   *// evaluating subset* X'_{opt} *by A learning algorithm*
If $(\delta > \delta_{opt})$
$X'_{opt} = X_{opt}$;
$\delta_{opt} = \delta$;
else
break and ***return*** X'_{opt}
end
return X'_{opt};
end

Figure II-8 A general explanation for embedded algorithm (Kumar & Minz, 2014)

Moreover, in embedded techniques the relationship between input feature and output feature is considered but it also searches locally for features that allow better local discrimination.

In summary, above two sections handled the way to select sensors. Every method has its own pros and cons in terms of application. In terms of selecting original sensors, which is original variables, sensor selection using variable subset section methods looks like more simple and intrinsic methods. However, in sensor selection using variable subset selection methods, its algorithm is applying classification algorithm before fault diagnosis and prognosis. It is kinds of process that conducting simple version of classification before classification. Due to the fact, still it takes high computational cost so that it might be hard to apply to real-time monitoring, where computational efficiency is essential.

Also, above two methods discussed are not the only methods to sensor selection. Having thought that factors influencing classification accuracy is input variable, Mahesh Pal uses Support Vector

Machine to feature selection in order to increase classification accuracy (Pal & Foody, 2010). The main purpose of methods discussed in previous section is also increasing classification accuracy.

There remain additional things to be discussed when selecting sensors apart from the selection method. Another key issue when selecting a subset of variables are relevance and redundancy. Criteria are different from researchers. Some researchers prefer to remove redundant sensors from the sensor list to improve the prediction accuracy. While, other researchers are cautious to remove redundant sensors because the removal of the redundant sensors may disregard the potential relevant sensors (Zhao *et al.*, 2010).

In conclusion, while dimensionality reduction algorithms such as PCA and LDA do well on sets of correlated features, sensor selection methods using variable subset selection perform poorly. They even fail to pick relevant variables, because the score they assign to correlated features is too similar, and none of the variables is strongly preferred over another. Hence, variable selection and dimensionality reduction algorithms have complementary advantages and disadvantages. Due to this complementary advantages and disadvantages, Janecek and Gansterer analyzed the relationship feature selection and classification accuracy (Janecek *et al.*, 2008). Subsets of the original variables are constructed using feature selection methods such as filter and wrapper techniques but also using PCA. Authors compared the classification results using three methods. The most remarkable insight of this study is that the principal components captured by the variance are not necessarily key indicators for the classification performance.

Thus, it is required to find out vital indicators among original variables increasing classification performance such as accuracy and computational efficiency.

2.2 Characterization of sensor signals for fault diagnosis

Figure II-9 shows typical representation of signals in time-domain. There are some measures to figure out central tendency of signal or change of signal such as mean, variance and so on. In fault diagnosis using univariate or multivariate Statistical Process Control (SPC), fault is usually detected focusing on the change in mean value or variance based on mahalanobis distance. Thus, finding out key characteristics of signals is critically related to an efficiency of fault diagnosis and prognosis performance.

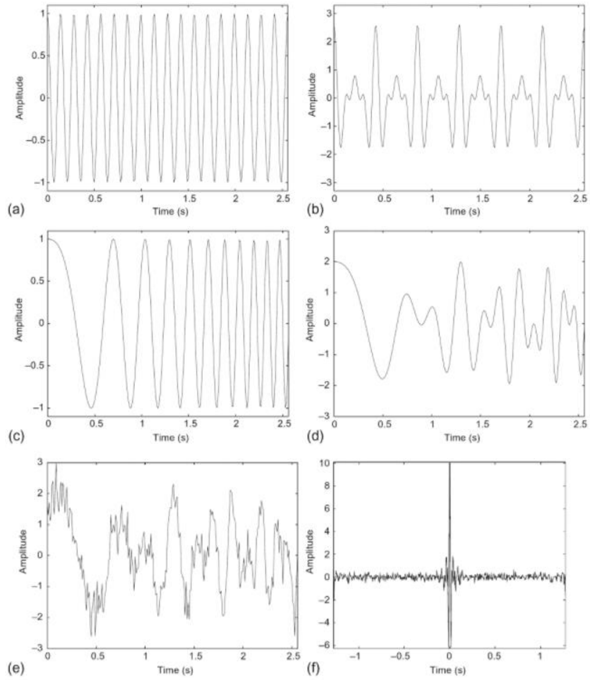


Figure II-9 Time--domain representations of the six signals (a) sinusoidal signal. (b) sum of sinusoids (c) monocomponent, nonstationary signal (d) multicomponent, nonstationary signal (d) sinc pulse with additive noise (Boashash, 2015)

For machine fault diagnostic, many kinds of researches have been done specifically to improve efficiency of fault diagnosis. Some researchers tried to extract features by analyzing signal characteristics which is meaningful to classify fault state in order to increase classification performance.

Statistical signal characterization was used to classify modulation signals (Hossen *et al.*, 2007). Author used statistical signal characterization for parameter extraction. Total four parameters are the amplitude mean, the period mean, the amplitude mean deviation and the period mean deviation which were extracted from the amplitude, frequency and phase of the signal waveform. Then, extracted four statistical signal characterization parameters were used to classification models as key features. Main objective of this study was simplifying the neural network process by using the small number of features.

Analyzing characteristics of signal was also used to finding effective features for blind digital modulation classification (Ebrahimzadeh & Ghazalian, 2011). The instantaneous characteristics, the higher order moments and the higher order cumulants were used as effective features in this study. Instantaneous characteristics consist of instantaneous phase or instantaneous frequency. Here, standard deviation of the absolute value of normalized and centered instantaneous frequency was used as an Instantaneous feature. These prominent characteristics of the signals helped discriminating digital signals. Another study has been done using high order cumulants to improve classification. Fourth-order cumulants was used as a one kind of characteristics for classifying various digital signaling formats (Swami & Sadler, 2000). It is simple so that it can work like a preliminary classifier. Similarly, the characteristics of sensor signals such as the kurtosis, the number of peaks in the phase probability density function, and the mean of the absolute value frequency were used as a key features (Lopatka & Pedzisz, 2000).

Other researchers used widely known characteristics of signal without proposing a new key characteristic. Root mean square, short time fourier transform, and characteristic frequency-band analysis were used to extract meaningful features from original dataset (Wu *et al.*, 2006).

As an another way to preprocess data before classification, unique characteristics of high-dimensional data were discussed in terms of geometrical and statistical properties (Jimenez & Landgrebe, 1998). Usually dimension reduction for supervised classification is conducted by computing data of full dimension. In this paper, preprocessing method considering characteristics of high-dimensional dataset was needed and parametric projection pursuit is the way to reduce the dimension by conducting calculation in the lower dimensional subspace. It is regarded as a preprocessing step for a feature extraction or classification steps. Overall steps are shown in the Figure II-10.

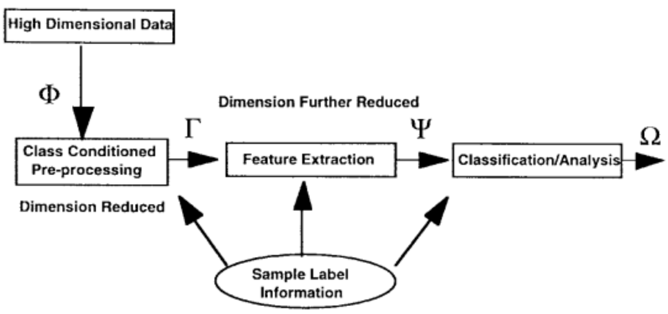


Figure II-10 Reprocessed high-dimensional data

Haining Liu proposed adaptive feature extraction method for machinery fault diagnosis using sparse coding as shown in the Figure II-11. A dictionary learning process was used, not manually defining basis functions. A dictionary learning process is useful to find the statistical structures of the signals. Then captured basis functions are used for sparse coding (Liu *et al.*, 2011).

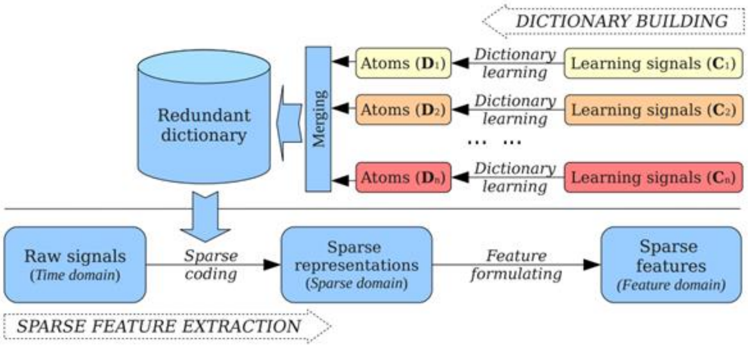


Figure II-11 The adaptive feature extraction scheme for machinery fault diagnosis based on sparse coding (Liu *et al.*, 2011)

To predict fault of equipment, Hu and Guo used fault tendency prediction based on multi-source information fusion (Hu *et al.*, 2012). They focused on the mutual relationship of fusion information source and the way to predict fault trend. In the fault tendency prediction model, information fusion is three-level including data level fusion, feature level fusion and decision level fusion. Especially in feature level fusion, the characteristic of feature is the characteristic of target equipment fault rule. In the process, extracted failure feature is correlated using a fusion method.

Another effort to enhance efficiency of fault diagnosis and prognosis is to analyze fault tendency. Wei Nai applied Naïve Bayesian Classifying to evaluate fault tendency percentage (Nai *et al.*, 2015). Naïve Bayesian Classifying method is rarely used in fault prognosis whereas it is widely used in fault diagnosis. Thus, author improved it to be adjusted to prognosis. Wang applied wavelet packet sample entropy to forecast fault trend of rolling element bearing (F. Wang *et al.*, 2011). He used Empirical Mode Decomposition to extract the signal trend.

To clearly figure out the change of signals or extract meaningful features for classification, preprocessing could be done before feature extraction such as filtering, noise cleaning and so on. Bugharbee used signal pretreatment before subjecting data to autoregressive model in fault diagnosis of rolling element bearings. Author highlighted steps of noise cleaning and stationarisation before autoregressive modelling. Singular spectrum analysis was used to clean noises. The proposed pretreatment process improved model prediction performance (Al-Bugharbee & Trendafilova, 2016).

2.3 Summary

In fault diagnosis and prognosis of machinery or systems, it is not always efficient to use all signal data. It is quite controversial issue amongst researchers whether classification performs better with reduced data or not. In a point of view that reduced data is helpful for classification, redundant data often cause difficulty in signal interpretation (Peter *et al.*, 2004). Thus, an efficient dimension reduction technique is necessary specifically for classification and sensor selection using dimension reduction techniques were discussed in previous section.

In summary, sensor selection can be done by space transformation and can be done by variable subset selection methods. Each method has its own characteristics. Using sensor selection method in transformed space, it is often hard to interpret newly made axes so instead we can use coefficient to select sensors among original sensors apart from original purpose of the techniques. Different from this method, sensor selection by variable subset selection is fundamentally to find best sensor subset. However, in this case some methods such as filter methods lack of considering dependency and interaction in change for computational efficiency. Otherwise, some methods considered dependency and interaction but instead it takes time to compute. Thus, the needs for computationally efficient and simple sensor selection methods arose. According to these needs, signal characteristics of sensor will be discussed. Analyzing and finding key characteristics of sensor signal is important to classification. Basic idea of this paper is that by analyzing key characteristics of sensor signals specifically aiming for classification, sensor can be ranked. Thus, for the first step, several key characteristics of signal are discussed and the way to select sensors based on the defined indices are proposed in the Chapter 3.

This Page Intentionally Left Blank

III. Characterization of sensor signals

3.1 Problem statement

The number of sensors monitoring the system increased as machinery system become complicated. It leads to multivariate analysis rather than simple univariate analysis in fault diagnosis and prognosis cause in complex system, correlation and interaction of several sensors have meaning in machinery failure. Thus, high dimensional data such as time series data should be processed.

When processing high dimensional sensor data, it is hard to use all sensors in analysis in practice. For example, in order to monitor vehicle engine fault forty sensors are used. However, if fault diagnosis is done with fault pattern analysis it takes tremendous time to derive fault pattern. Also, redundancy should be considered. Still it is controversial issue for the fundamental needs of dimension reduction. It is not always the case that rich information can tell explainable. In some cases, too many sensor data might be redundant. Redundant data have potential to lower prediction accuracy.

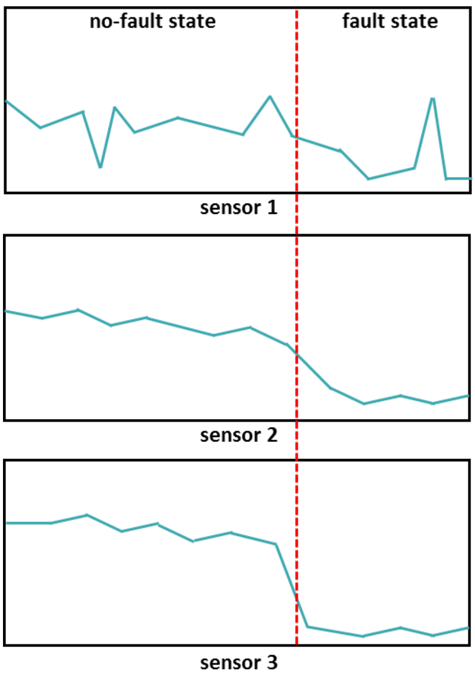


Figure III-1 Examples of similar signal trends

As shown in the Figure III-1, if signal trends of sensor2 and sensor3 are similar, which means highly correlated, one sensor can be removed in analysis. Then, instead of using all sensors, reduced sensors can be used for classification reducing redundancy effect. Thus, properly reducing dimension of data is necessary.

In case of dimension reduction, several things should be in consideration: first, the number of reduced dimension, second, the way to reduce dimension. For the number of sensors selected, sometimes it fully depends on the users or in other case, it can be set using threshold or using automatically defined algorithms. As discussed before, rationale for reducing dimension should be valid, which means assumptions should be carefully in consideration. In some case, the assumption is that preferring high dimensional data, without loss of information whereas in other case, to reduce redundant information, dimension reduction is proposed for preprocessing step before fault detection.

Also, existing dimension reduction techniques has several limitations especially when selecting original variables (sensors) specific for fault detection. Thus, to solve these problems, new sensor selection method using key characteristics of sensor signals will be presented.

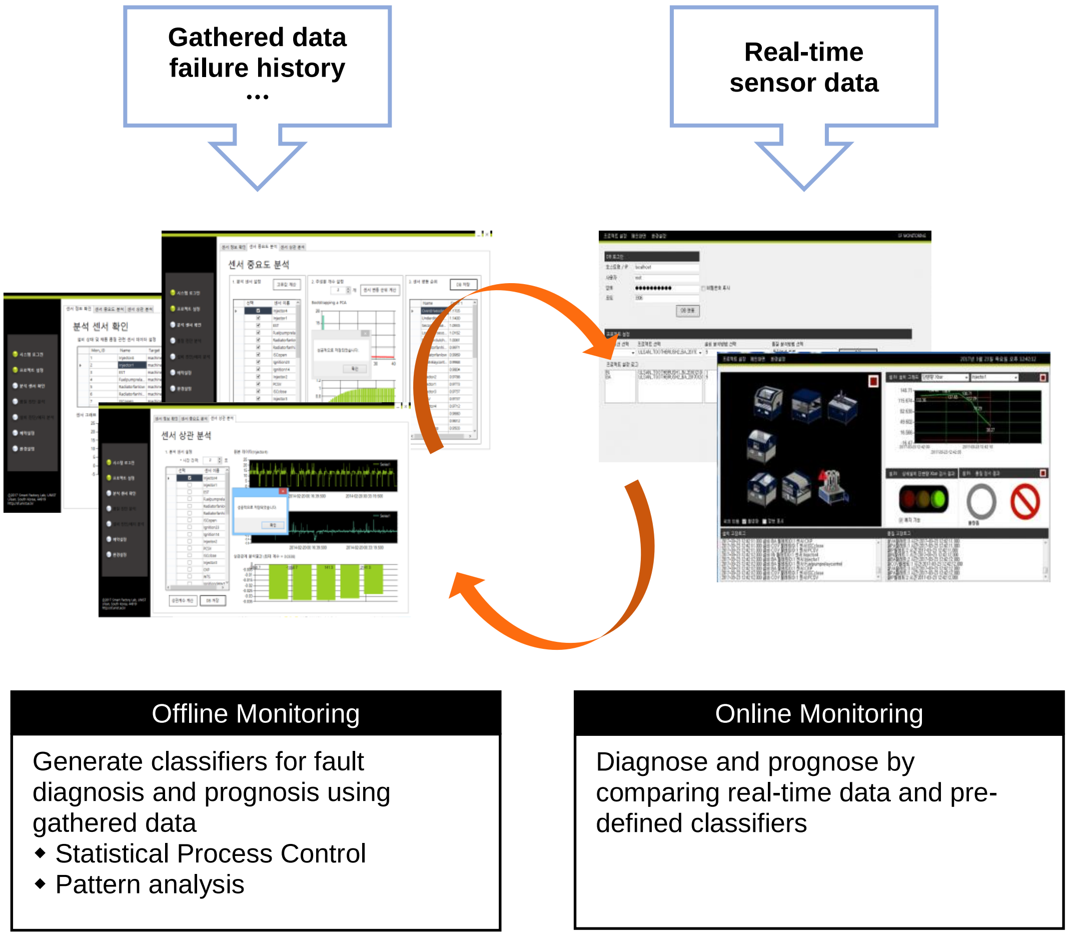


Figure III-2 Overall framework for real-time fault diagnosis

3.2 Characteristics of sensor signals

In order to identify significance of specific sensor, characteristics of the sensor signal can be examined. Variance, mean, kurtosis are measures to examine overall trend or helpful indicators for comparison. Each method is simple so that easy to compute but cannot always cover whole characteristics of signal.

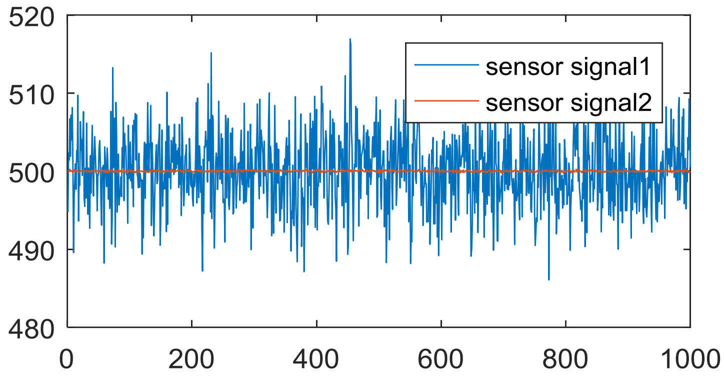


Figure III-3 An example of equal mean sensor signals

For example, when using mean to compare signals, changes in signal are ignored. As shown in the Figure III-3, even though sensor1 is in unstable state whereas sensor2 is in stable state, mean value of both sensors are same so that mean value cannot distinguish these two sensors. Also, when using variance usually it assumes that a sensor which has big variance can have much more informative data than a sensor having small variance.

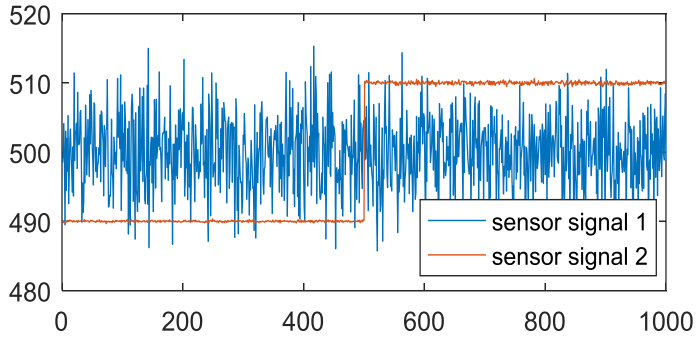


Figure III-4 An example of equal variance sensor signals

In this case, abrupt changes could be ignored as shown in the Figure III-4. Even though sensor1 has repeated vibration whereas sensor2 has only one drop, two signals have equal variance so that it is hard to distinguish them using variance. Thus, it is necessary to have new measures to classify fault and no-fault state effectively. In this regard, in next section new measures for sensor signals will be discussed.

3.3 New measures for sensor signals

There are mainly three indices will be discussed in this section: (i) abrupt variance (ii) discernibility index and (iii) sparse impulse. Abrupt variance and discernibility index were discussed and derived in (Baek & Kim, 2017). Newly proposed index is a sparse impulse index which can compensate the things which abrupt variance and discernibility index cannot cover.

3.3.1 Abrupt variance (*aVar*)

Abrupt variance is derived from variance. As mentioned above, it is not always the case that huge total variance of data cannot ensure it to be a vital indicator for classification.

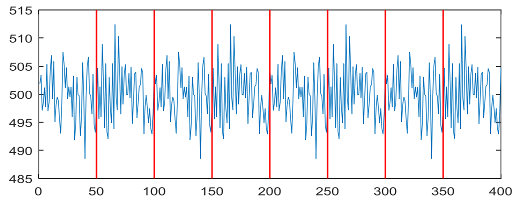


Figure III-5 A signal which cannot differentiate fault and no-fault state. Vertical line means state change; normal to fault state or fault to normal state

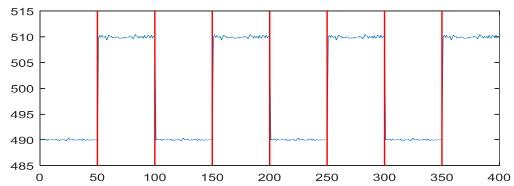


Figure III-6 A signal which can differentiate fault and no-fault state. Vertical line means state change; normal to fault state or fault to normal state

For example, as shown in the Figure III-5 and the Figure III-6, total variations are similar but detail signal changes differ from states. Like the case, overall variance sometimes neglects detail information such as a change of signal which can be meaningful to classification. Thus, it is necessary to have a measure which can distinguish fault and no-fault state well even with the small variance. Also, the highly fluctuated sensor signals mean that system is in unstable state and under this circumstance, an abrupt variance (*aVar*) index was developed (Baek & Kim, 2017).

$$\boldsymbol{aVar_i} = \frac{\sum_{j=1}(x_{ij}-\bar{x}_i)^2}{n} \times \frac{\sum_{j=1}((x_{ij+1}-x_{ij})-(\overline{x_{ij+1}-x_{ij}}))^2}{n-1}$$

where

x_{ij+1} the j^{th} data of the i^{th} sensor

$\bar{x}_{ij}$ the mean of the j^{th} data of the i^{th} sensor

$\overline{x_{ij+1}-x_{ij}}$ the mean of the difference of the i^{th} sensor

This abrupt variance index is available in both time series data and frequency data simultaneously.

3.3.2 Discernibility index (*DI*)

For the second index which reflects key characteristics of sensor signals is discernibility index (*DI*) (Baek & Kim, 2017).

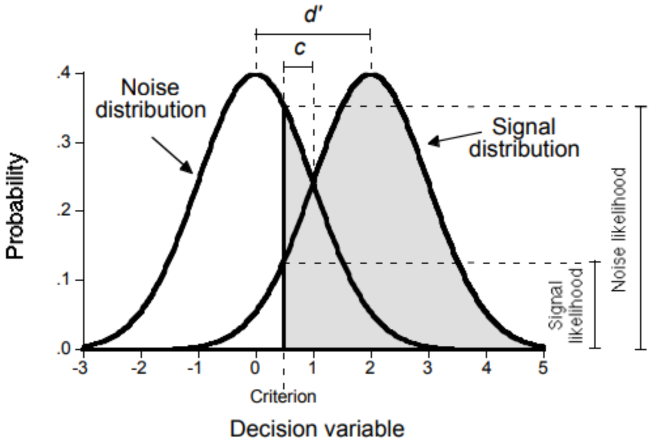


Figure III-7 Distribution of the decision variable across noise and signal. *d'* is the sensitivity index (Stanislaw & Todorov, 1999)

In signal detection theory, there already exists sensitivity index d'. It measures the separation between the means of the signal and the noise distributions. Sensitivity index can be drawn with following equation:

$$\mathrm{d}' = \frac{\mu_S - \mu_N}{\sqrt{\frac{1}{2}(\sigma_S^2 - \sigma_N^2)}}$$

where,

μ_S mean of the signal

μ_N mean of the noise

σ_S standard deviation of signal

σ_N standard deviation of noise

Important assumption is that two distributions are normally distributed. However, in real application it is hard to meet the assumption. Thus, some researchers have used nonparametric measures to derive sensitivity. Pollack and Norman proposed A' which is widely known measure among several nonparametric measures of sensitivity (Pollack & Norman, 1964). Similarly, A discernibility index is an area measure, which is derived with following equation.

$$DI_i = \int \min\{PDF^n, PDF^f\}\,dx$$

where,

PDF^n the estimated probability density function of sensor data in normal state

PDF^f the estimated probability density function of sensor data in fault state

For clear comparison between sensors, normalization for *DI* should be done.

$$DI(\mathrm{X}) = \mathrm{Norm}\left(\frac{1}{m}\sum_{i=1}^{m} DI_i\right)$$

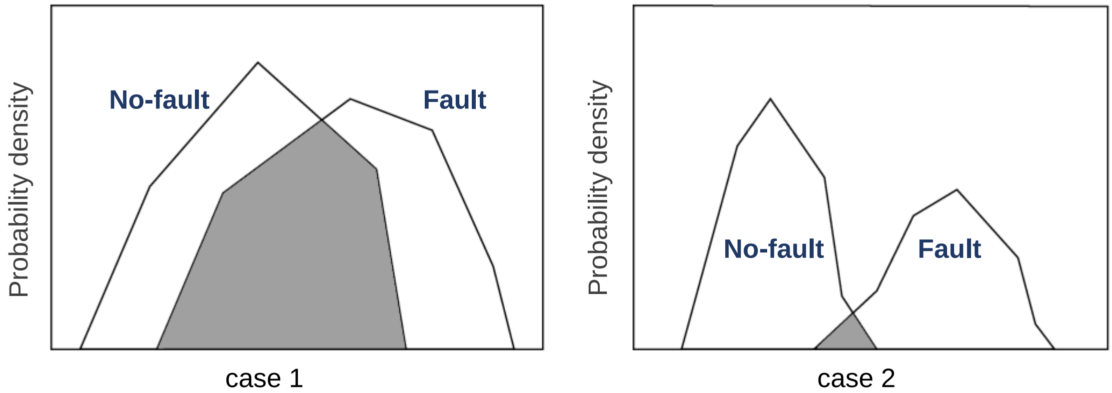


Figure III-8 (a) high DI case (b) low DI case

As clearly shown in the Figure III-8, two states are highly discernible which means separable when having low *DI* value than having high *DI* value.

3.3.3 Sparse impulse (*SI*)

Generally, sensors which have high variance are selected in many dimension reduction techniques such as PCA and PLS. Basic assumption of it is that the sensor which has high variance can have change which is much explainable for original data than the sensor which has low variance. However, there might be a case that even a sensor having neglectable variance is effective having useful information to classify fault and no-fault state.

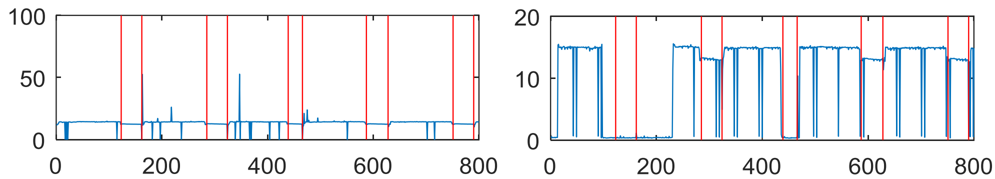


Figure III-9 Examples of low and high variance sensor signals

For example, in the case shown in the left in the Figure III-9, the sensor is powerful to classify fault and no-fault state even though it has comparatively low variance whereas in the case shown in the right in the Figure III-9, even the sensor has high variance it does not show clear difference between different states.

Sparse impulse (*SI*) is derived based on these assumptions. It detects sparsely existing impulse signals and considers its effect to distinguish states.

Here, impulse signal is defined as a sudden peak. It is shown that sparse impulses are detected in the Figure III-10 (left). There is an important constraint for sparse impulse, which is minimum height (H_{min}). Not every peak is regarded as sparse impulse. Peaks which exceed N times standard deviation of the total signal are counted as sparse impulses (here, N is set as three). Since overall data are high dimensional time series, peaks are counted in specific time window segment, which is user defined (here, time window segment is 10). There are mainly two strategies in terms of defining sparse impulse. In first strategy, if there exists at least one peak value in a time window segment, it is counted as a one sparse impulse as shown in the Figure III-10 (center). In second strategy, mean value is considered as a reference for sparse impulse. If mean value in a specific time window segment is larger than threshold, it can be counted as a one sparse impulse in the Figure III-10 (right). User can decide which methods to select for sparse impulse considering overall sensor signal conditions.

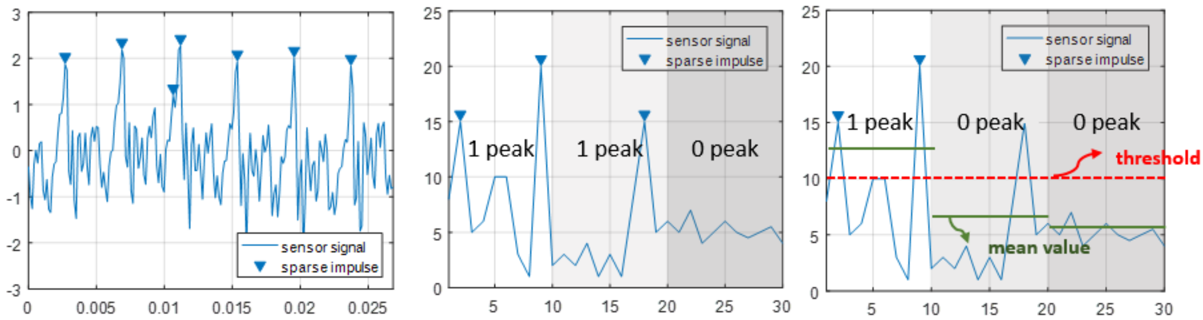


Figure III-10 Sparse impulse signal examples

Sparse impulse score is derived with the number of impulses. For clear understanding, we can represent an impulse signal of which direction is up as a positive impulse (PI) and an impulse signal of which direction is down as a negative impulse (NI).

$$SI = \frac{\omega a + 1}{\omega a + 1 + b}$$

where,

ω is the user-defined constant which can adjust the weight for the score

a is the difference of the number of impulses between fault and no-fault sections

b is the number of impulses which considers compensatory effects of impulses with opposite direction *a* and *b* values especially differ by existence of positive and negative impulses. Thus, it is calculated case by case (i) only positive impulse or negative impulse exists and (ii) both positive and negative impulse exist. It can be explained more detail by several examples in the Figure III-2. It is mainly divided into two cases. First, if only positive impulse or negative impulse exists, it gives highest score for the case where impulse exists only in one state whereas if impulse exists in both cases the difference of the number of impulses between each state is also considered for scoring. Second, if both PI and NI exists, sparse impulse in opposite direction have compensatory effect. In the left down example of the Figure III-2, even though the number of sparse impulse is equal for both states, the directions of sparse impulses are opposite so that does not deduct score. For more detail explanation is listed as a pseudo code in the Table III-1.

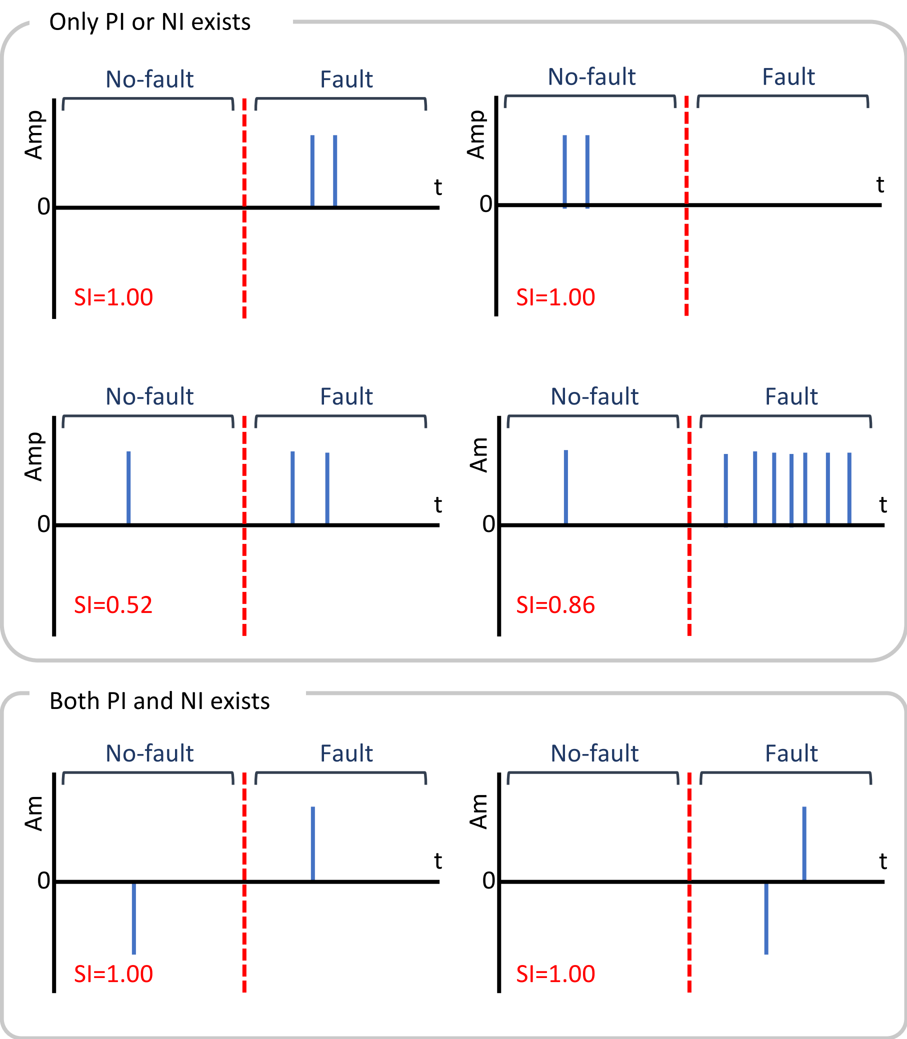


Figure III-11 Examples of spare impulse signals

Table III-1 Pseudo code of sparse impulse score

Algorithm: Sparse impulse score

Require:

```
PI_n(f) = the number of positive impulse in normal(fault) state
NI_n(f)= the number of negative impulse in normal(fault) state
diff_PI = PI_f-PI_n
diff_NI = NI_f-NI_n
ω=weight for the score
for i=1:the number of samples
Case 1)Both PI and NI exist
        if diff_PI~=0
            score1=(ω*diff_PI+1)/( ω*diff_PI+1+min(PI_f,PI_n))
        end
        if diff_NI~=0
            score2=(ω*diff_NI+1)/(ω*diff_NI+1+min(NI_f,NI_n))
        end
        if diff_PI~=0 && diff_NI~=0
            score= (score1+score2)/2
        elseif diff_PI~=0 && diff_NI==0
            score= score1
        elseif diff_PI==0 && diff_NI~=0
            score= score2
        end
Case 2-1) only NI exists
        alpha= diff_NI
        if alpha<=0
            beta=NI_f
            if beta==0
                score=1
            else
                score=(w*abs(alpha)+1)/(w*abs(alpha)+1+beta)
            end
        else
            beta=NI_n
            if beta==0
                score=1
            else
                score=(ω*alpha+1)/(ω*alpha+1+beta)
            end
Case 2-2) only PI exists
        alpha= diff_PI
        if alpha<=0
            beta=PI_f
            if beta==0
                score=1
            else
                score=(w*abs(alpha)+1)/(w*abs(alpha)+1+beta)
            end
        else
            beta=PI_n
            if beta==0
                score=1
            else
                score=(ω*alpha+1)/(ω*alpha+1+beta)
            end
        end
```

It seems that total three indices are defined for index reflecting key characteristics related to fault detection. However, those indices can be calculated in both time domain and frequency domain so that overall six indices exist.

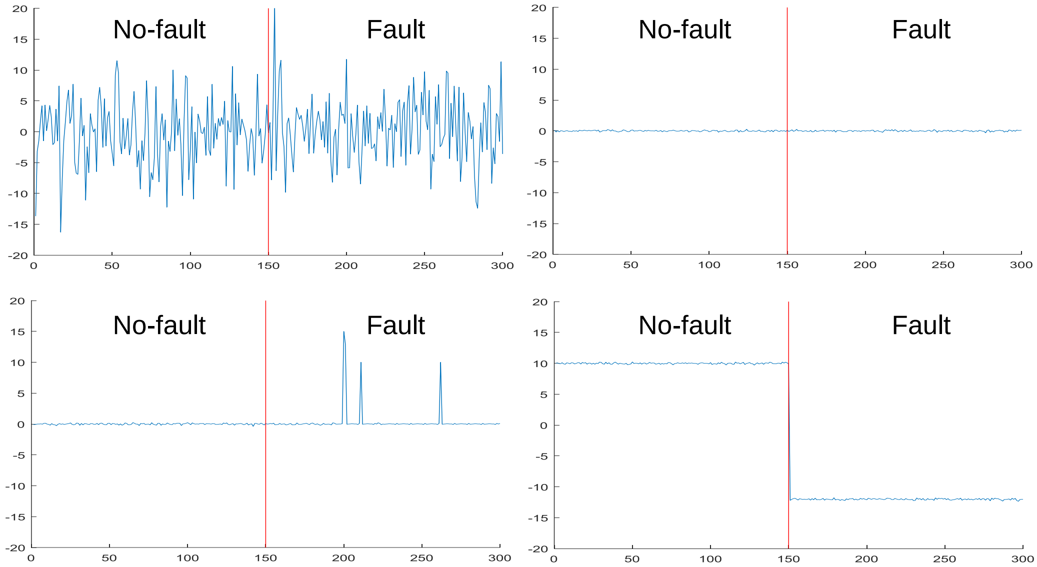


Figure III-12 Example signals considering *aVar* and *DI*. In clockwise (a-d). (a): high *aVar* and high *DI* (b): low *aVar* and high *DI* (c): low *aVar* and low *DI* (d): high *aVar* and low *DI*

Using the indices, even we can infer the shape of sensor signals. Detail examples are in the Figure III-12 and the Figure III-13. In the Figure III-12, there are example signals only considering *aVar* and *DI*. Combining these two indices, we can even infer the signal shape using signal characteristics. Also, example signals considering *aVar* and *SI* are shown in the Figure III-13.

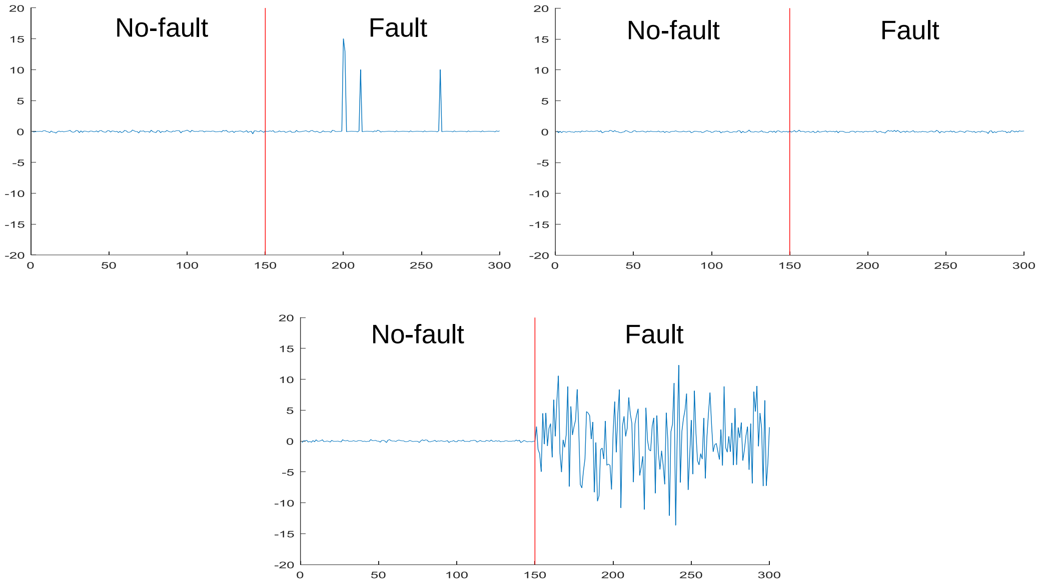


Figure III-13 Example signals considering *aVar* and *SI*. In clockwise (a-c). (a): high *aVar* and high *SI* (b): low *aVar* and low *SI* (c): high *aVar* and low *SI*.

3.4 Sensor selection methods

For the next step of discussing the key characteristics of sensor signals, a procedure to select sensors should be discussed. In this section, mainly three methods are proposed. First, sensors are selected based on abrupt variance. In this method, sensors which have low abrupt variance are selected as suitable sensors for fault detection. Second, all indices such as abrupt variance, discernibility index, sparse impulse are used to select sensors.

3.4.1 *aVar*-based PCA

A first method for sensor selection is based on abrupt variance, which was discussed before. Abrupt variance can be used as two versions in sensor selection.

Simplified model: abrupt variance and variance

It is already defined that low abrupt variance means that signal changes steadily in time series. Signals can be classified with variance and *aVar* into big four categories like the Figure III-14.

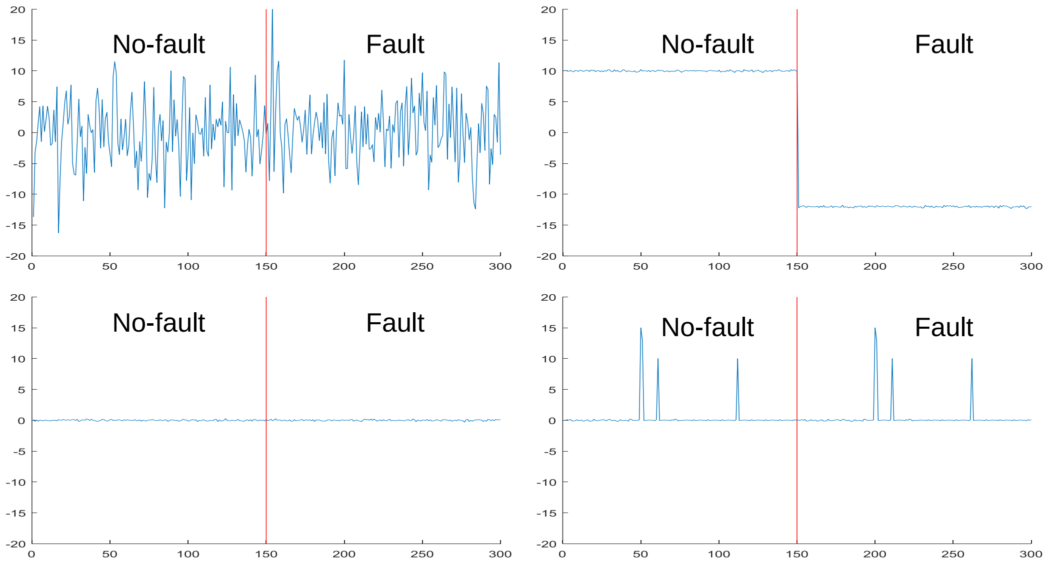


Figure III-14 Classified four signals. Clockwise (a-d) (a): high variance and high *aVar* (b): high variance and low *aVar* (c): low variance and low *aVar* (d): low variance and high *aVar*

Among those classified signals, signal which is helpful to classify fault and no-fault state is the signal (b) which have high variance and low $aVar$. Thus, maximization function can be derived as follows:

$$\underset{sensor_i}{argmax}(var_i - aVar_i)$$

Based on *aVar* definition, formation of it is a multiplication of a constant α and variance. Thus, above equation become simplified like:

$$\underset{sensor_i}{argmax}(var_i(1-\alpha))$$

With above simple maximizing function, sensors could be ranked.

Extended PCA model: *aVar*-based PCA

Principal Component Analysis is one of the widely used methods for dimension reduction. In the procedure of PCA, principal components are drawn, which can express overall data in reduced space. Considering the very basic equation of PCA, fundamental principle of PCA is maximizing variance. By the definition, aVar is modified version of variance. Thus, basic assumption is as follows:

Assumption: *The lowest N principle components which are drawn with aVar-PCA can be used to select sensors having low abrupt variance*

To select sensors using aVar-based PCA, there are mainly four steps needed.

STEP1. Calculate abrupt variance of dataset
STEP2. Calculate covariance using aVar
STEP3. Conduct Principal Component Analysis using covariance in STEP2
STEP4. Bottom N (user defined) sensors are selected based on the coefficient matrix of aVar-based PCA

For the first step, abrupt variance of dataset is calculated. After that, Covariance of abrupt variance should be calculated. Covariance can be drawn using the relations between variates like the equation below:

$$var(X+Y) = var(X) + var(Y) + 2cov(X,Y)$$

$$cov(X,Y) = \frac{var(X+Y) - var(X) - var(Y)}{2}$$

Using the equation, covariance of abrupt variance can be calculated. Next step is conducting PCA. PCA can be done by singular vector decomposition of data matrix, or by eigenvalue decomposition of data covariance or correlation matrix. In this paper, the latter method is used. Detail steps and explanations for conducting PCA using eigenvalue decomposition are as follows:

x is N dimensional original dataset. Principal components are derived by maximizing the variance of the projected data on the reduced dimension or minimizing the mean squared distance between the data and projected data. In this paper, we handled first approach, maximizing variance.

Variance of the projected data is abbreviated to $var(x)$

$$var(x) = \frac{1}{N}\sum_{n=1}^{N}(u^T x_n - u^T \bar{x})^2 = u^T S u$$

where, S is the covariance of the data matrix and u is a unit vector on which data are projected.

$$\mathrm{S} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \bar{x})(x_n - \bar{x})^T$$

Using Lagrange multiplier, maximization function is formulated

$$\mathrm{L}(\mathrm{x}, \lambda) = u^T S u + \lambda(1 - u^T u)$$

Derivative $\mathrm{L}(\mathrm{x}, \lambda)$ with regard to u

$$Su = \lambda u$$

where,
u is an eigenvector of covariance matrix and λ is an eigenvalue of covariance matrix.

By multiplying u^T, the equation above is simplified.

$$u^T \mathrm{S} u = \lambda u^T u$$

$$u^T \mathrm{S} u = \lambda$$

Thus, maximization variance is done by finding maximum eigenvalue in this regard. *aVar* can be represented as multiplication of variance and a constant. Thus, in the same way, *aVar*-based PCA is also done with covariance of abrupt variance. Coefficient matrix which is the result of *aVar*-PCA is used to select sensors which is significant for classifying fault and no-fault state. More detail formula of *aVar* is as follows.

Abrupt variance of the projected data is abbreviated to $aVar(x)$

$$aVar(x) = \frac{1}{N}\sum_{n=1}^{N}(u^T x_n - u^T \bar{x})^2 \times \frac{1}{N-1}\sum_{n=1}^{N-1}(u^T {x^{diff}}_n - u^T \overline{x^{diff}})^2 = u^T \mathrm{S} u u^T S^{diff} u$$

where,

x is the original data matrix

x^{diff} is the difference of original data matrix in time series. (e.g., $x_{i,j+1} - x_{i,j}$ means the difference between (j+1)th data and jth data of ith sensor)

S is the covariance of x

S^{diff} is the covariance of x^{diff}

u is a unit vector on which data are projected

Since u is a unit vector, $uu^T = 1$, the equation is simplified.

$$u^T S u u^T S^{diff} u = u^T S S^{diff} u$$

Using Lagrange multiplier, maximization function is formulated

$$\mathrm{L}(\mathrm{x}, \lambda) = u^T \mathrm{S} S^{diff} u + \lambda(1 - u^T u)$$

Derivative $\mathrm{L}(\mathrm{x}, \lambda)$ with regard to u

$$\mathrm{S} S^{diff} u = \lambda u$$

By multiplying u_1^T, the equation above is simplified.

$$u^T \mathrm{S} S^{diff} u = \lambda u^T u$$

$$u^T \mathrm{S} S^{diff} u = \lambda$$

Thus, in the same way to PCA, *aVar* maximization (or minimization) problem turned into maximization (or minimization) of eigenvalue. Basic assumption is same with previous method. The number of selected sensors is user-defined value (N). For the next step, sum coefficients from first to N in coefficient matrix and rank the summed value. Finally, N sensors with lowest summed values are selected according to the procedure.

Above procedure are to find sensors, which are original variables of dataset. Aside from this, if it is not the case using original variables, new principal component based on abrupt variance could be proposed to be used in fault detection.

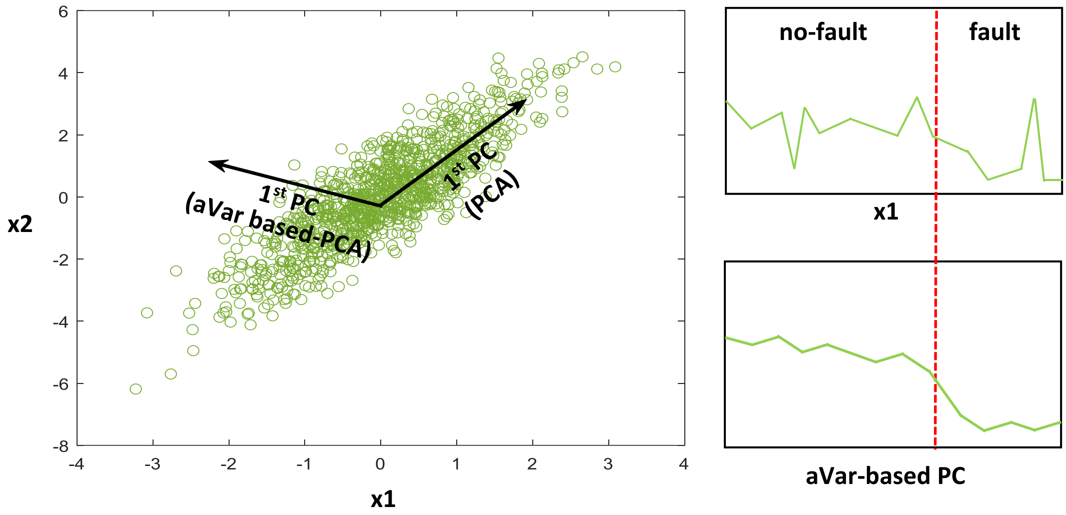


Figure III-15 The first principal component from abrupt variance-based PCA and the first principal component from conventional PCA (left) and the expected result (right)

Considering Original PCA, data which are projected onto principal components can represent original dataset well but it does not guarantee that fault detection performance increases in projected space. Compared to original PCA, if data are projected onto principal components based on abrupt variance, converted data are much more sensitive to abrupt variation. In other words, projected data are in the space which supports classification of fault and no-fault state.

3.4.2 Weighted sum approach

A Second method for sensor selection is done by aggregating PCA, *aVar* and *DI*. Basically, sensors are ranked with weighted sum of each index.

Not only aVar, DI but also PCA are included for sensor ranking procedure since PCA is the most powerful dimension reduction methods. Total three indices are derived from each method: ind_{PCA}, ind_{aVar}, and ind_{DI}. ind_{PCA} is derived by sum of PCA coefficient. As defined previous sections 3.4.1, *aVar*-based PCA is done with applying *aVar* to PCA so ind_{aVar} is derived by inverse of sum of aVar-based PCA coefficient since small *aVar* of a specific sensor means the signal characteristic of that sensor is good for distinguishing fault and no-fault state. Lastly, ind_{DI} is derived by inverse of *DI* since low discernibility index of a sensor means the signal of the sensor is highly discernible. Using these three indices aggregated index is derived as follows:

$$\text{Aggregated index} = ind_{PCA} + ind_{aVar} + ind_{DI}$$

In this way, all sensors can be ranked. Additionally, *SI* is used to finding sensors which were low-ranked in PCA sensor selection but might have potential to be used to distinguish system state. *SI* score of low-ranked N sensors in PCA sensor selection method is calculated and sensors which exceed threshold (e.g., 0.7, 0.8) are recommended to be used to fault diagnosis additionally.

In this way, total rank of sensors is attained. In this stage, I let users decide the number of sensors selected. When reducing dimensions of data, information loss occurs. In other words, it means it would be the best case if it is possible to use the sensors as many as possible. Thus, here the number of sensors are user-defined value (N). Overall framework is described in the Figure III-16.

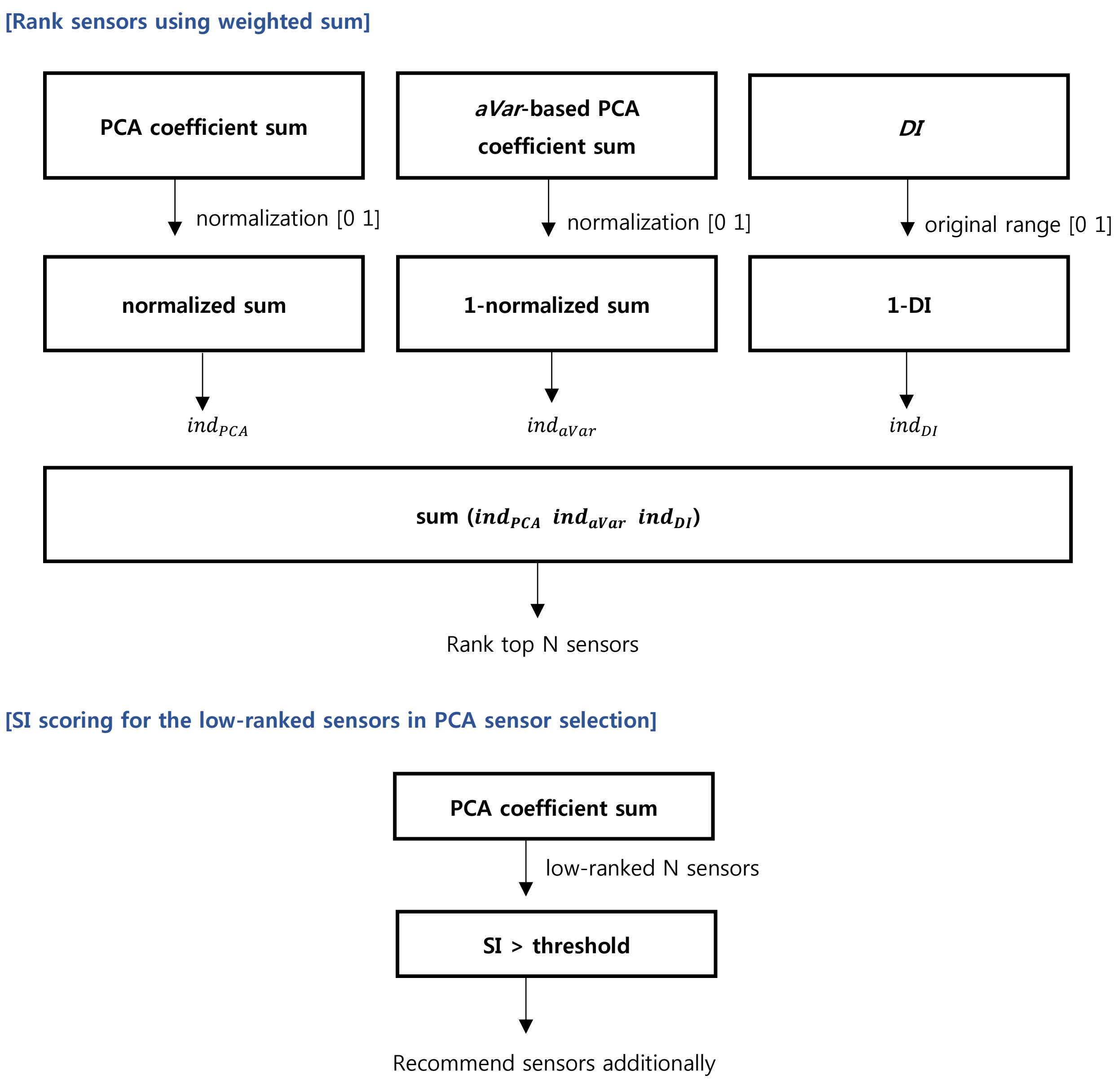


Figure III-16 Overall framework for weighted sum approach sensor selection

IV. Case study

A series of experiments have been conducted to validate the proposed sensor selection method of this thesis.

4.1 Vehicle diagnostics simulator

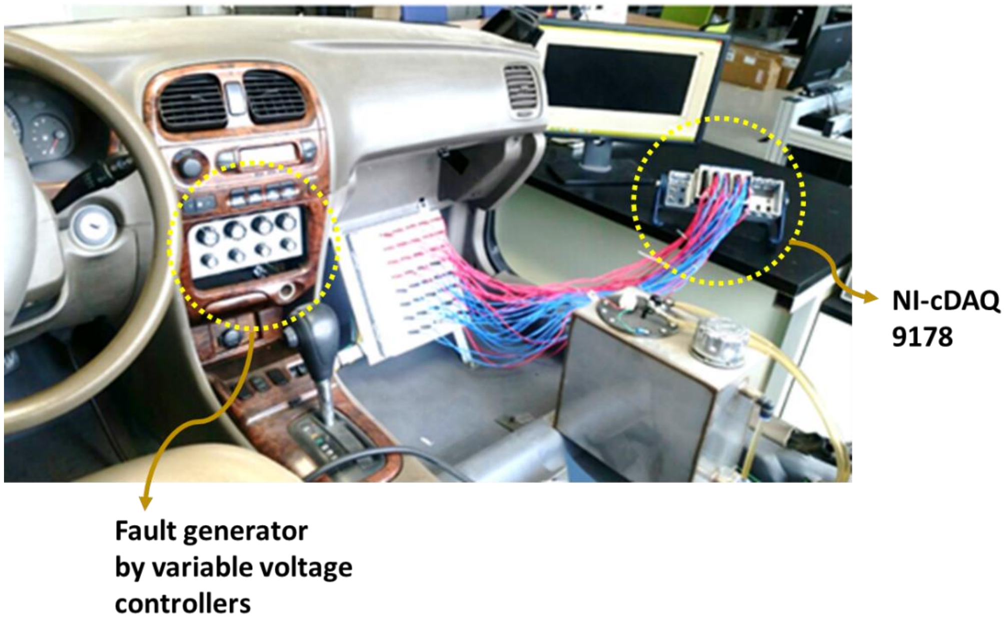


Figure IV-1 Vehicle diagnostics simulator

Sensor data were collected from vehicle diagnostics simulator in the smart factory laboratory as shown in the Figure IV-1. Fault state defined in this experiment was knocking in the vehicle engine. More detail information for the experimental setting is as follows.

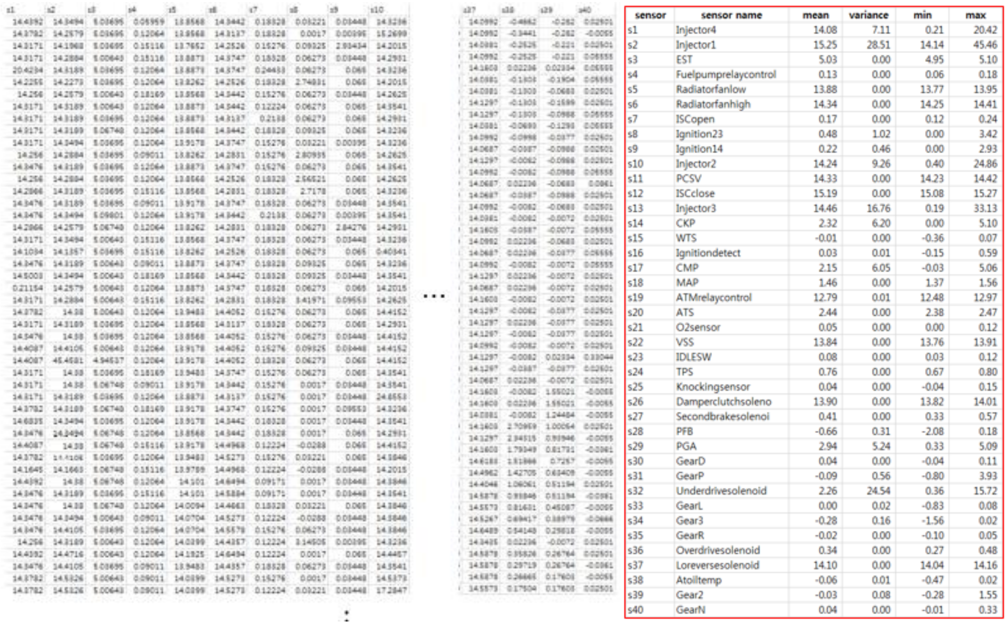
4.1.1 Experimental setting

Table IV-1 Description of the experimental data using vehicle fault generator

Fault state definition	Engine knocking and abnormal engine RPM
Fault generating method	Control intake air pressure randomly Control actuators in the fuel injection system

The number of sensors	Total forty sensors
Type of sensors	Injector, crank position, manifold absolute pressure (MAP), throttle position (TP) …
Sampling rate	2 Hz

Since total number of sensors are forty, dataset contains tremendous number of signals in high dimension Figure IV-2.



sensor	sensor name	mean	variance	min	max
s1	Injector4	14.08	7.11	0.21	20.42
s2	Injector1	15.25	28.51	14.14	45.46
s3	EST	5.03	0.00	4.95	5.10
s4	Fuelpumprelaycontrol	0.13	0.00	0.06	0.18
s5	Radiatorfanlow	13.88	0.00	13.77	13.95
s6	Radiatorfanhigh	14.34	0.00	14.25	14.41
s7	ISCopen	0.17	0.00	0.12	0.24
s8	Ignition23	0.48	1.02	0.00	3.42
s9	Ignition14	0.22	0.46	0.00	2.93
s10	Injector2	14.24	9.26	0.40	24.86
s11	PCSV	14.33	0.00	14.23	14.42
s12	ISCclose	15.19	0.00	15.08	15.27
s13	Injector3	14.46	16.76	0.19	33.13
s14	CKP	2.32	6.20	0.00	5.10
s15	WTS	-0.01	0.00	-0.36	0.07
s16	Ignitiondetect	0.03	0.01	-0.15	0.59
s17	CMP	2.15	6.05	-0.03	5.06
s18	MAP	1.46	0.00	1.37	1.56
s19	ATMrelaycontrol	12.79	0.01	12.48	12.97
s20	ATS	2.44	0.00	2.38	2.47
s21	O2sensor	0.05	0.00	0.00	0.12
s22	VSS	13.84	0.00	13.76	13.91
s23	IDLESW	0.08	0.00	0.03	0.12
s24	TPS	0.76	0.00	0.67	0.80
s25	Knockingsensor	0.04	0.00	-0.04	0.15
s26	Damperclutchsoleno	13.90	0.00	13.82	14.01
s27	Secondbrakesolenoi	0.41	0.00	0.33	0.57
s28	PFB	-0.66	0.31	-2.08	0.18
s29	PGA	2.94	5.24	0.33	5.09
s30	GearD	0.04	0.00	-0.04	0.11
s31	GearP	-0.09	0.56	-0.80	3.93
s32	Underdrivesolenoid	2.26	24.54	0.36	15.72
s33	GearL	0.00	0.02	-0.83	0.08
s34	Gear3	-0.28	0.16	-1.56	0.02
s35	GearR	-0.02	0.00	-0.10	0.05
s36	Overdrivesolenoid	0.34	0.00	0.27	0.48
s37	Loreversesolenoid	14.10	0.00	14.04	14.16
s38	Atoiltemp	-0.06	0.01	-0.47	0.02
s39	Gear2	-0.03	0.08	-0.28	1.55
s40	GearN	0.04	0.00	-0.01	0.33

Figure IV-2 High dimensional data. Total number of sensors is 40

In order to select sensors, methods which were discussed in the Chapter 3 will be all used. (i) sensor selection using weighted sum of indices (ii) sensor selection using *aVar*-based PCA.

To validate the proposed sensor selection method, Hotelling T^2 was used. Model for test was constructed with normal dataset. PCA model, abrupt variance-based PCA model, and aggregated index -based model will be constructed and then, using test dataset performance of normal and fault classification will be evaluated regarding hit rate and false alarm rate.

In Hotelling T^2 test, threshold to decide whether state is in normal or in fault is c.

$$c = \frac{(N-1)}{(N-p)} F_{p,N-p}$$

where, N is the number of sample and p is the number of principal components

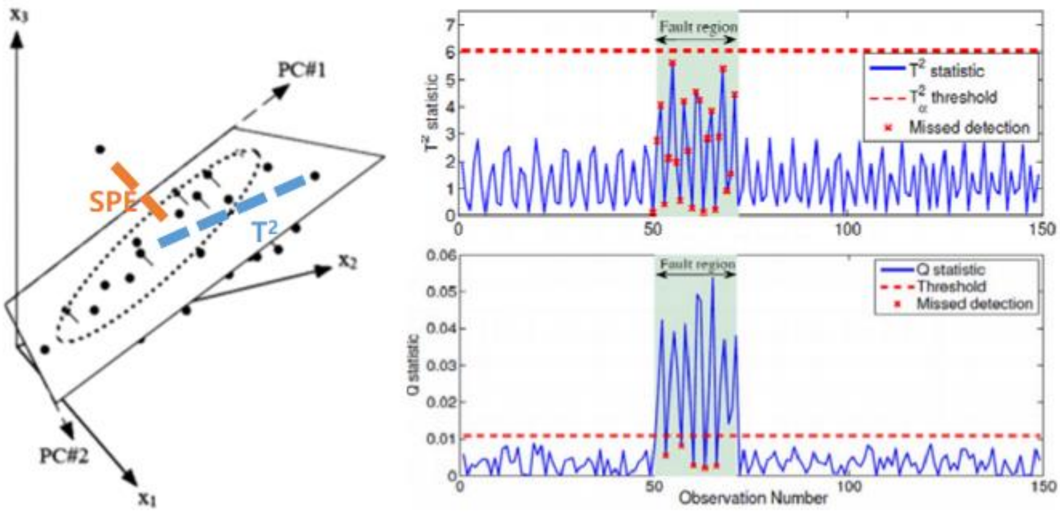


Figure IV-3 Hotelling T^2 and Q statistics (Baek *et al.*, 2016)

Hypotheses for detection is as follows

H0: signal absent (no-fault state)

H1: signal present (fault state)

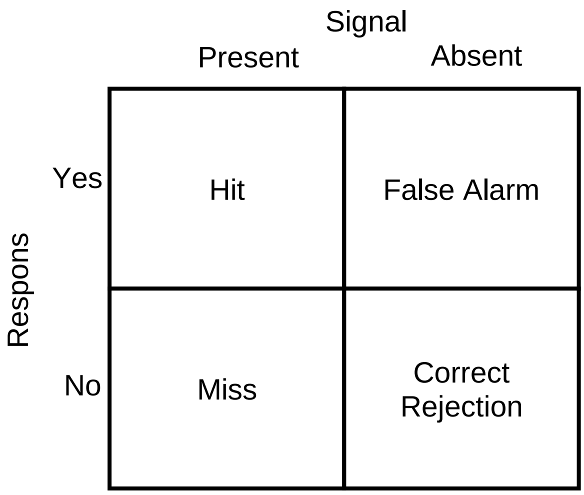


Figure IV-4 Signal detection theory

In signal detection theory, trials are sorted into four categories such as hit, false alarm, miss, and correct rejection. Among four cases shown in the Figure IV-4, criteria for each case differs from industrial application. For instance, in semiconductor industry, missing faults of wafer results critical impact on the quality of final semiconductor. Also, in the power plant industry, missing machinery faults or facility faults might result untold losses. Thus, in the fields where even very slight faults or defects are not allowed, missing rate is one of the most considerable factors. However, for detection performance, hit rate is most important, which catch fault in time. Considering those characteristics, we can evaluate performance of detection algorithms.

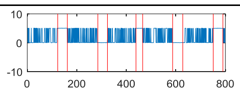
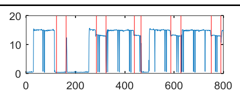
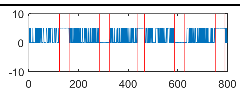
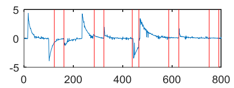
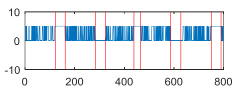
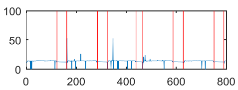
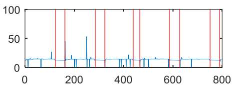
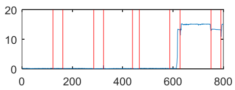
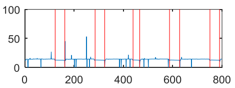
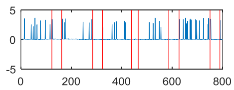
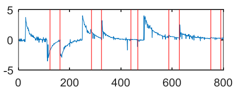
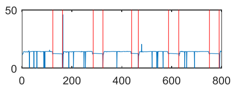
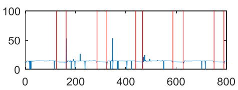
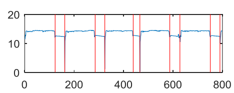
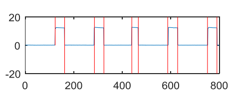
4.1.2 Experimental results and discussion

Sensor selection methods used in this experiment are (i) sensor selection using original PCA, (ii) sensor selection using *aVar*-based PCA and (iii) sensor selection using aggregated index.

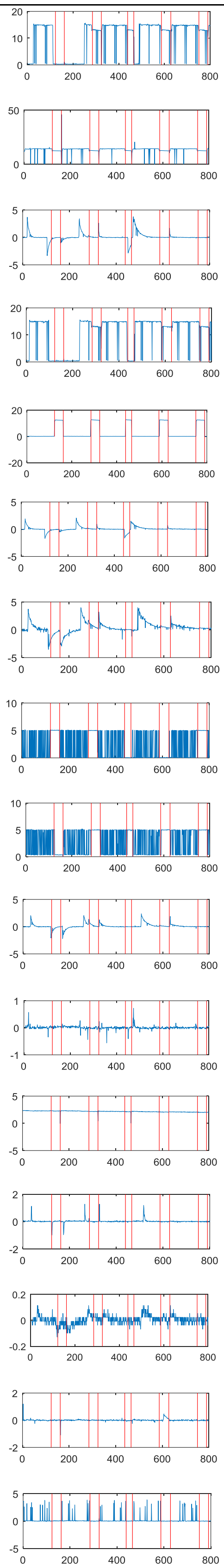
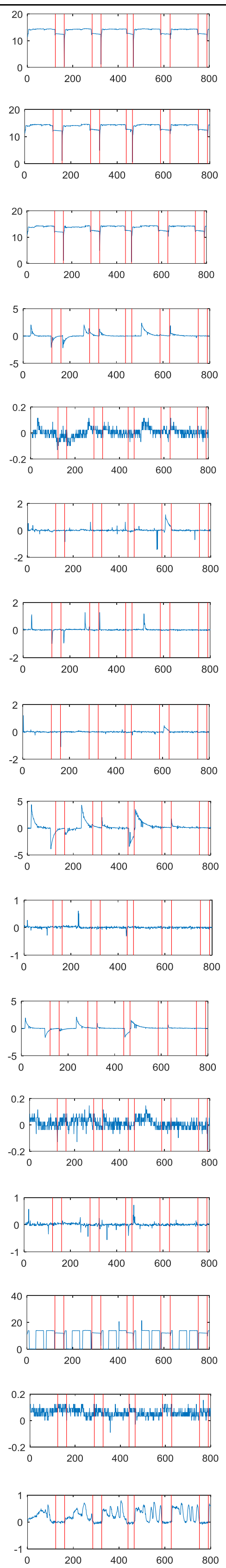
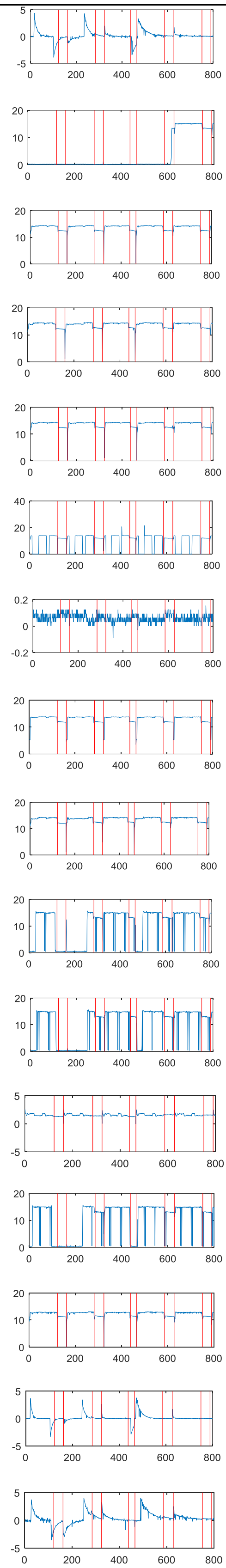
For clear understanding, we attached sensor signal plots starting from highest rank (see the Table IV-2). As clearly shown in the, PCA selects sensors which has large variance and there are no any other criteria. Thus, the sensor rank does not show any trend or tendency except large variance. Also, it does not show clear distinguishable signal characteristics among top ranked sensors.

Different from the PCA sensor selection results, several sensors such as sensor31, sensor8, sensor36 ranked lower in aVar-based PCA sensor selection. Those sensors ranked high when using PCA sensor selection just because they have large variance. However, several sensors such as sensor27, sensor12, sensor28 ranked higher than before with consideration of abrupt variance. In *aVar*-based PCA sensor selection, sensors which have unfluctuating signals with low abrupt variance highly-ranked. Also, several sensors such as sensor12, sensor6, sensor37 are top-ranked when using aggregated index based sensor selection, which ranked lower when using PCA sensor selection. Sensors with high *SI* score which was low-ranked in PCA sensor selection was highlighted with bold in the Table IV-2.

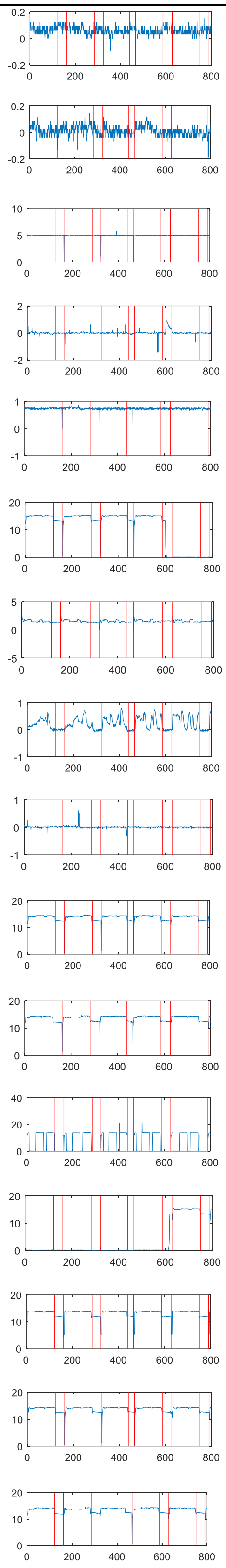
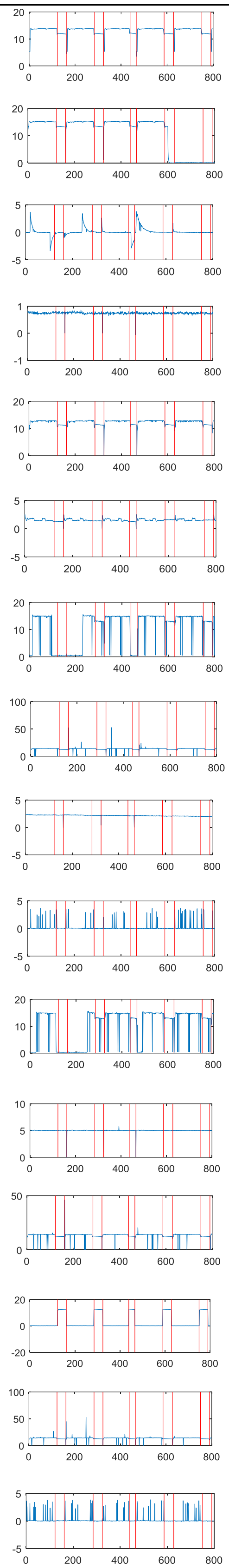
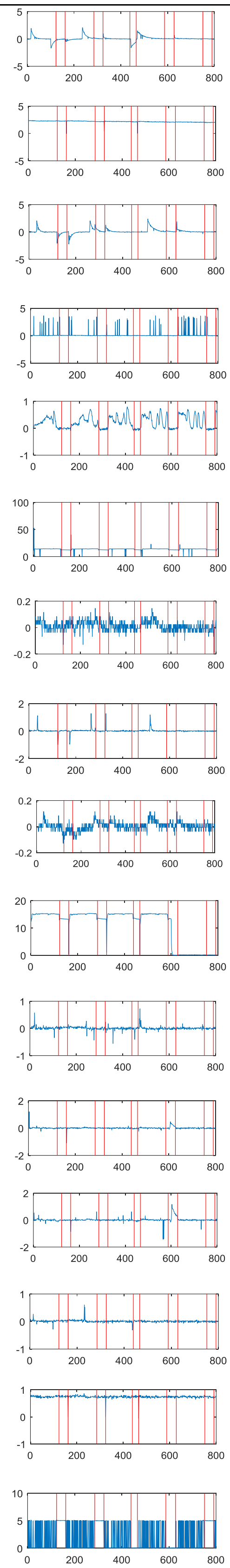
Table IV-2 Sensor selection results (using 15 sensors)

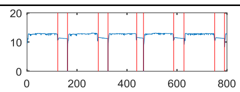
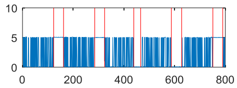
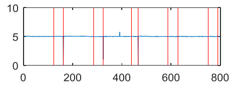
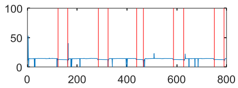
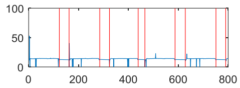
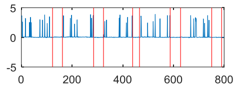
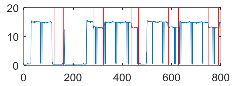
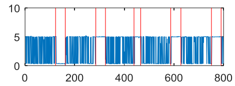
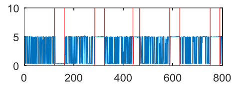
Rank	#	PCA sensor selection	# (PCA rank)	aVar-based PCA sensor selection	# (PCA rank)	Aggregated index sensor selection
1	17		27 (40)		17 (1)	
2	31		17 (1)		1 (5)	
3	2		12 (34)		2 (3)	
4	8		28 (12)		10 (7)	
5	1		11 (36)		4 (10)	

Rank	#	PCA sensor selection	# (PCA rank)	aVar-based PCA sensor selection	# (PCA rank)	Aggregated index sensor selection
6	36		6 (31)		31 (2)	
7	10		37 (32)		**12 (34)**	
8	38		26 (37)		**6 (31)**	
9	32		34 (15)		**37 (32)**	
10	4		35 (19)		**11 (36)**	
11	39		16 (25)		**5 (33)**	
12	28		33 (18)		23 (22)	
13	14		15 (20)		**22 (35)**	
14	29		31 (2)		**26 (37)**	
15	34		40 (30)		27 (40)	
16	30		39 (11)		36 (6)	
17	20		25 (23)		**18 (28)**	
18	33		30 (16)		32 (9)	
19	35		5 (33)		**19 (38)**	
20	15		23 (22)		38 (8)	
21	9		21 (29)		28 (12)	

Rank	#	PCA sensor selection	# (PCA rank)	aVar-based PCA sensor selection	# (PCA rank)	Aggregated index sensor selection
22	23		22 (35)		39 (11)	
23	25		7 (27)		20 (17)	
24	3		38 (8)		34 (15)	
25	16		24 (26)		8 (4)	
26	24		19 (38)		**21 (29)**	
27	7		18 (28)		**13 (39)**	
28	18		32 (9)		25 (23)	
29	21		1 (5)		33 (18)	
30	40		20 (17)		35 (19)	
31	6		8 (4)		7 (27)	
32	37		36 (6)		30 (16)	
33	5		3 (24)		15 (20)	
34	12		10 (7)		16 (25)	
35	22		4 (10)		40 (30)	
36	11		2 (3)		24 (26)	
37	26		9 (21)		14 (13)	

Rank	#	PCA sensor selection	# (PCA rank)	aVar-based PCA sensor selection	# (PCA rank)	Aggregated index sensor selection
38	19		14 (13)		3 (24)	
39	13		13 (39)		9 (21)	
40	27		29 (14)		29 (14)	

As shown in the Figure IV-5, it shows better detection performance regarding hit and false alarm when using sensors selected with proposed method than using sensor selected with original PCA. Detection performance results are listed in the Table IV-3. In dimension reduction, the reduced dimension has critical effect on the classification result so the number of sensors used for detection differs from 5 to 15. PCA result shows slightly higher hit rate than proposed methods result in the case using 15 sensors whereas proposed methods result shows lower false alarm than the PCA result. As for 10 sensors and 5 sensors example, the number of sensors so small that it cannot generate proper PCA model which can explain original dataset so hit rate of each case is almost zero. It is shown that as the number of selected sensors decreases, proposed methods outperform than PCA sensors selection regarding hit rate.

However, several concerns arise. Regarding proposed methods, detection performance seems to be improved. However, proposed indices were made considering detection only. Thus, focus of key characteristics of sensor was not incipient fault but abrupt and obvious change in signals. Considering prognosis, which means prediction, further studies are necessary.

Table IV-3 Detection performance comparison

# of selected sensor / **Sensor selection method**	**15 sensors**		**10 sensors**		**5 sensors**	
	Hit (%)	**FA (%)**	**Hit (%)**	**FA (%)**	**Hit (%)**	**FA (%)**
PCA	99.27	24.46	0.03	0.57	0.00	0.00
aVar-based PCA	98.86	25.36	98.26	4.63	97.85	3.43
Aggregated index based	98.75	9.52	98.63	3.88	97.97	2.86

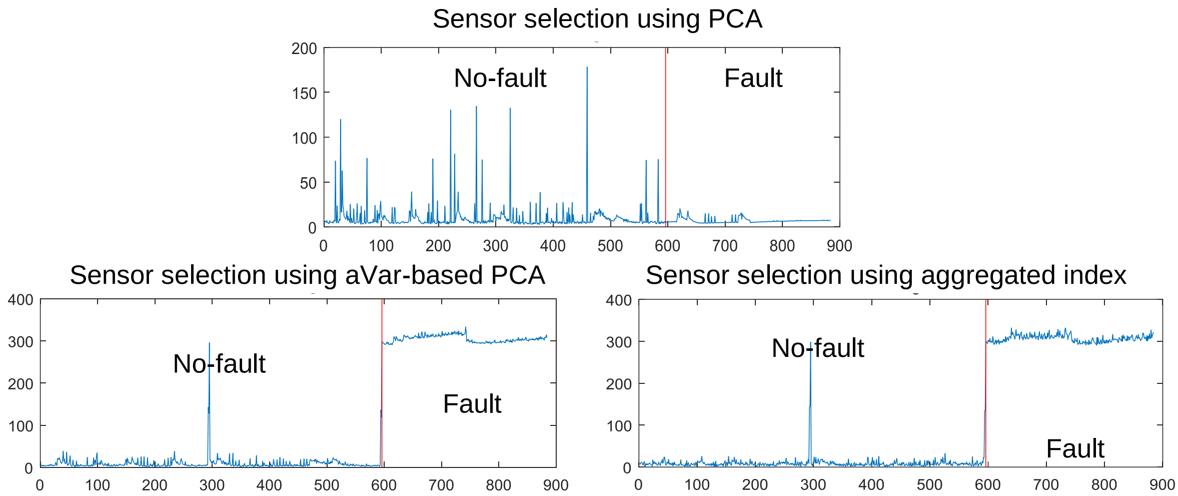


Figure IV-5 Plot Hotelling T^2 results to clearly show the difference (10 sensors)

4.2 Gear system diagnostics simulator

The proposed sensor selection methods were applied to gear diagnostics simulator, which uses vibration to system monitoring.

4.2.1 Experimental setting

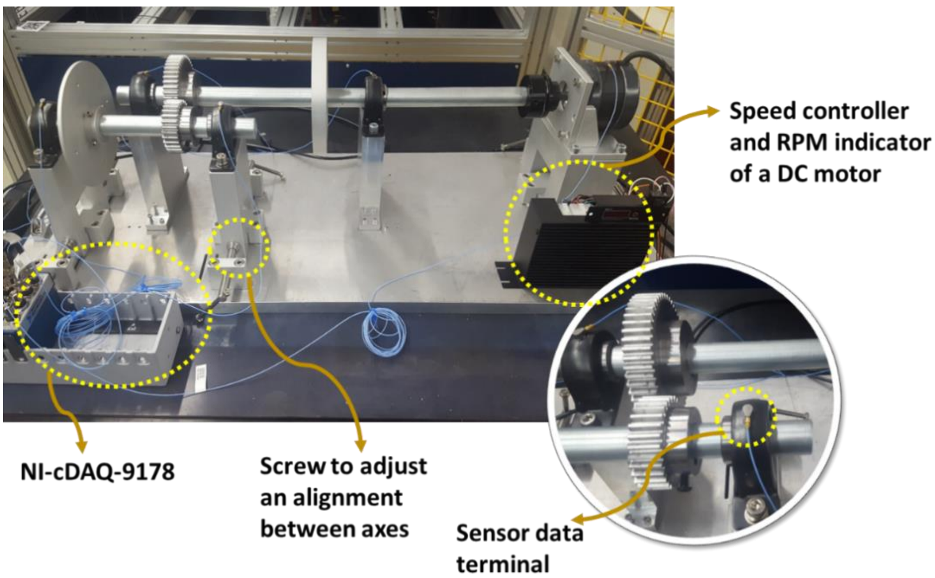


Figure IV-6 Gear system diagnostics simulator

To get sensor data which can be used to machinery monitoring, a gear diagnostics simulator was designed and made. More detail hardware specification of this simulator will be listed with figures and explanation in the Table IV-4. Conventional gear simulator consists of single axis only. However, I wanted to generate fault mode generating an axis unalignment so this simulator consists of two axes.

Table IV-4 Hardware specification of gear system diagnostics simulator

Part name	Part	Explanation
Gears		Two planetary gears connecting two axes
Bearing unit		Pillow-type bearing unit Ball bearing Total five units are attached in the pillows
Motor		BLDC motor MAX 3000 RPM
Motor driver		Motor driver for BLDC motor Motor speed, driving direction can be controlled
RPM meter		Motor RPM indicator
Rotors		For making unbalance to the axis, two rotors were customized to installing weights having four tabbed holes
Flexible coupling		Give flexibility to the long axis and the motor
Fixation screw and other parts		Tool for controlling the location and the angle of short axes Fixation screws and other fixing parts prevent pillow block from moving or vibrating
Accelerometer and accessories (e.g., cable)		single-axis PCB accelerometer sensitivity 10mV/g Total five sensors are attached on the bearing unit

To gather sensor data for system monitoring and validate performance of diagnostic and prognostics algorithm using the gathered, this simulator has several characteristics as follows.

1. Continuous and consistent data acquisition
2. Various failure mode
3. Easily change the system state from fault state to normal state, different from reliability analysis

Detail information of data selected and fault mode is listed in the Table IV-5

Table IV-5 Description of the experimental data using machinery gear fault simulator

Fault state definition	Abnormal vibration
Fault generating method	· Loosen the screw fixing bearing case · Misalignment · Mass unbalance of a rotor
Motor specification	Max 3000 RPM
The number of sensors	Total five sensors
Type of sensors	Accelerometer

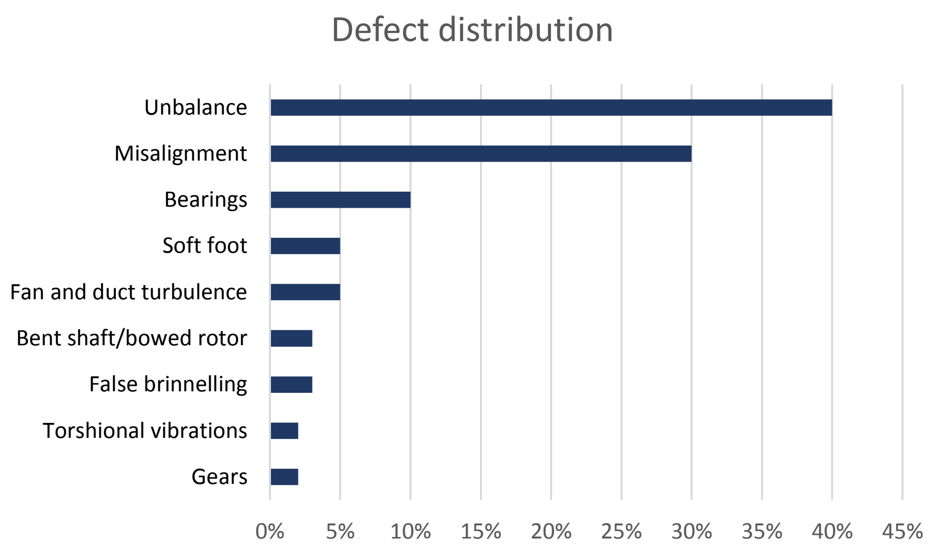


Figure IV-7 Defect distribution of rotating machinery (Wowk, 1991)

Fault modes are total three: (i) bearing fault (ii) misalignment and (iii) mass unbalance of rotor. Generally, unbalance, misalignment, and bearing faults comprise the great majority of rotating machinery faults as shown in the Figure IV-7. Thus, most of the machinery faults can be handled by monitoring those three kinds of fault modes. Fault conditions can be made by first, loosening the screw which fixes the bearing cases on the pillars, second making misalignment by adjusting gaps between pillars supporting axis and third, adding additional weights on the rotors. In reality, defects of bearings include mechanical damage, crack damage, wear damage, lubricant deficiency and so on (McInerny & Dai, 2003). However, it was hard to make artificial damage on the bearing in the experiment, so it was substituted with loosening bearing case.

Not only defining fault modes, sensor type for data acquisition should be determined. In this simulator, vibration sensors are attached which can cover most kinds of defects.

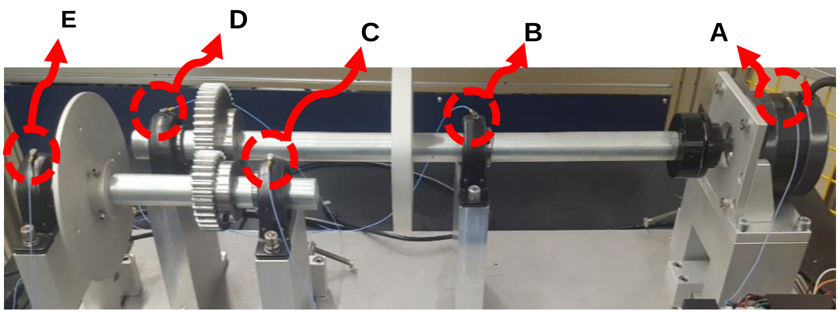


Figure IV-8 Sensor location and sensor numbering. A-E: accelerometer sensor 1-5

Five accelerometers are attached on the top of four bearing cases and a motor. Each sensor is located close to possible failure points.

Experiments were done in five different conditions: First, normal condition without any misalignment, unbalance, and bearing faults, second, misalignment of two axes, third, mass unbalance of the rotor and forth and last, loosen bearing case of different bearings.

4.2.2 Experimental results and discussion

Using the dataset produced by gear fault simulator, total four fault mode data are used by combining normal condition data with four fault condition data.

Fault mode 1: Misalignment of two axes, short axis is not in parallel with long axis
Fault mode 2: Mass unbalance of rotor in long axis
Fault mode 3: Loosen bearing case where sensor 2 is attached
Fault mode 4: Loosen bearing case where sensor 3 is attached

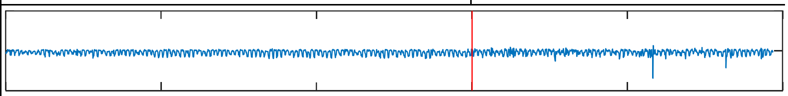
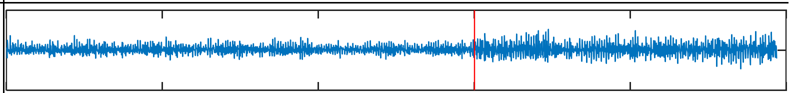
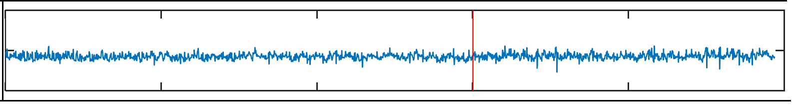
Methods used for sensor selection are (i) sensor selection using coefficients of original PCA, (ii) sensor selection using *aVar*-based PCA and (iii) sensor selection using aggregated index.

Table IV-6 Experimental results of each fault mode

`	sensor number	*aVar*	*Variance*	*SI*	*DI*
Fault mode 1	1	0.00004	**0.00426**	0.12841	0.20709
	2	0.00006	0.00553	0.21118	0.17985
	3	0.00047	0.01441	0.18619	0.10505
	4	1.51553	1.40842	0.22211	0.04773
	5	**4.36148**	1.94495	0.16375	0.02701
Fault mode 2	1	0.00002	**0.00268**	0.32621	0.19601
	2	0.00005	0.00498	0.32537	0.17478
	3	0.00016	0.00802	0.35765	0.13360
	4	0.85903	1.17548	0.38313	0.01466
	5	**3.07869**	1.79822	0.27610	0.01530
Fault mode 3	1	0.00002	**0.00306**	0.28023	0.19048
	2	0.00016	0.00880	0.21827	0.12013
	3	0.00019	0.00889	0.31367	0.13423
	4	0.84816	1.26790	0.26466	0.01862
	5	**4.11381**	2.41638	0.16845	0.01266
Fault mode 4	1	0.00002	**0.00292**	0.09475	0.27673
	2	0.00003	0.00357	0.30530	0.26686
	3	0.01447	0.07393	0.05530	0.08388
	4	0.89728	1.21255	0.11146	0.02063
	5	**4.93760**	2.40457	0.04928	0.02927

Experimental results such as *aVar*, variance, *SI*, and *DI* of each sensor are listed in the Table IV-6. In whole dataset, variance of sensor1 is the smallest, which is attached on the motor. Regarding *aVar*, sensor5 has the largest *aVar* value, which means it has highly fluctuating characteristic. For detail explanation, signal visualization, rank and other information of analyzed data are listed in the following tables (Table IV-7~Table IV-10).

Table IV-7 Visualization of selected sensors in fault mode 1

Sensor selection method	Sensor rank	Sensor number	Signal visualization (No-fault / Fault)
PCA	1	4	
	2	3	
	3	5	

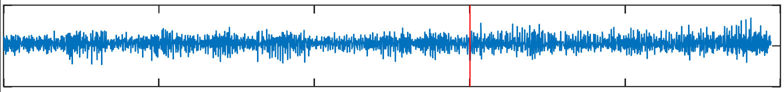
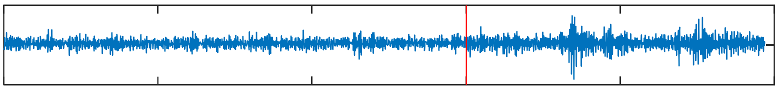
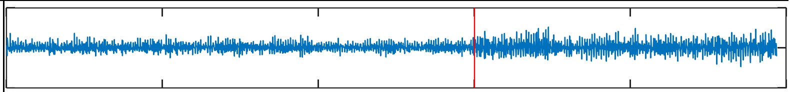
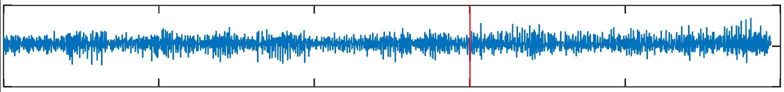
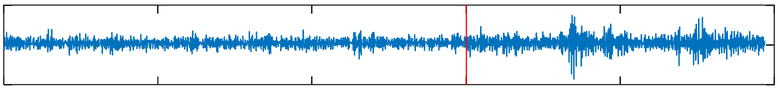
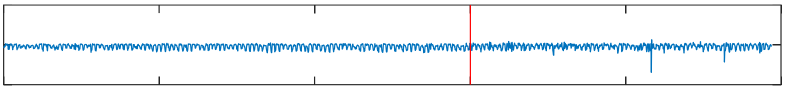
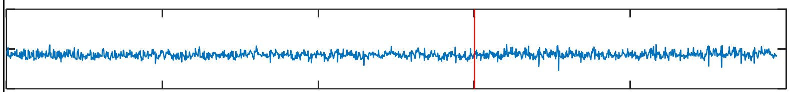
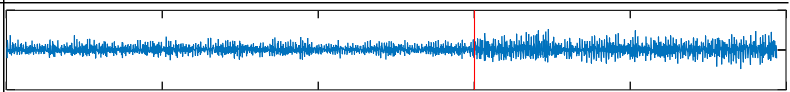
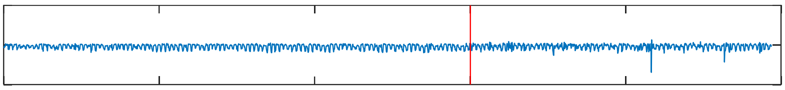
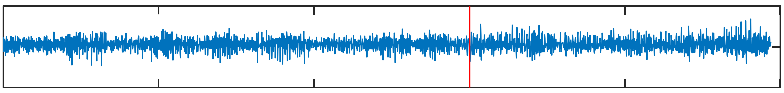
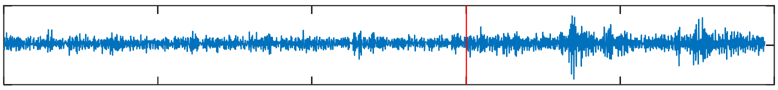
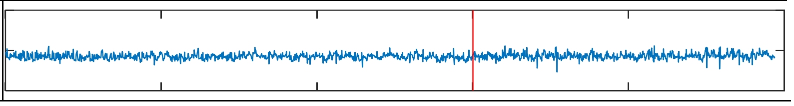
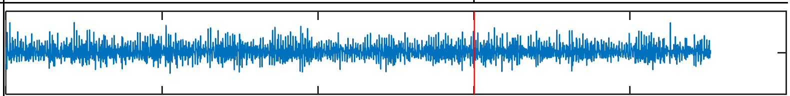
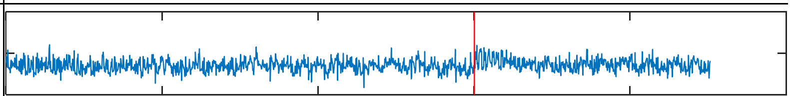
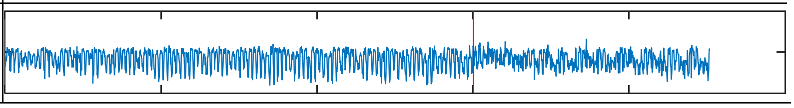
Sensor selection method	Sensor rank	Sensor number	Signal visualization	
			No-fault	Fault
	4	2		
	5	1		
aVar-based PCA	1	3		
	2	2		
	3	1		
	4	4		
	5	5		
Aggregated index	1	3		
	2	4		
	3	2		
	4	1		
	5	5		

Table IV-8 Visualization of selected sensors in fault mode 2

Sensor selection method	Sensor rank	Sensor number	Signal visualization	
			No-fault	Fault
PCA	1	4		
	2	3		
	3	5		

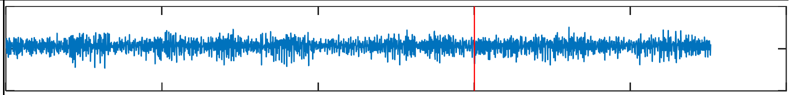
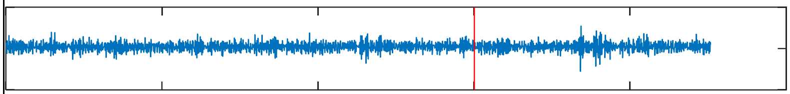
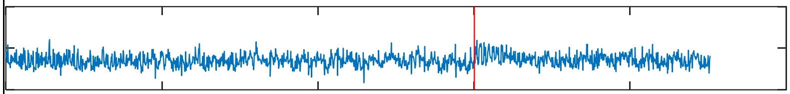
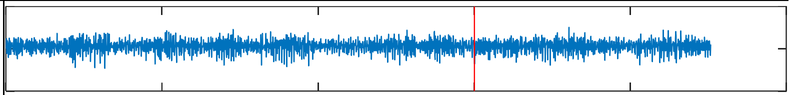
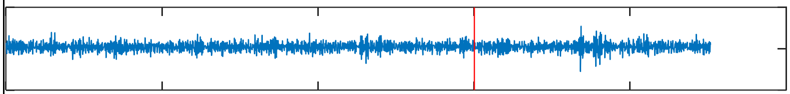
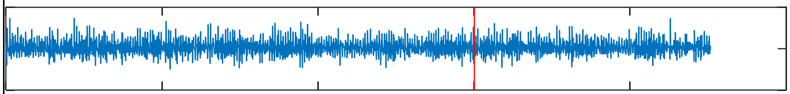
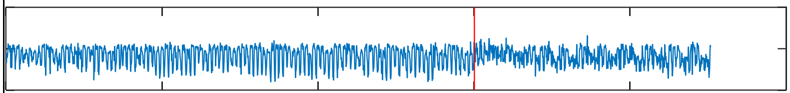
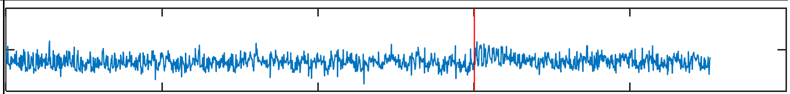
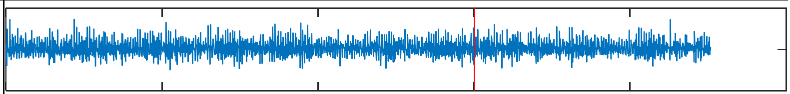
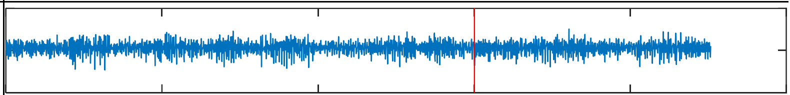
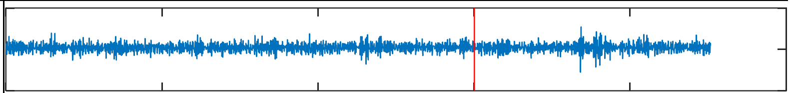
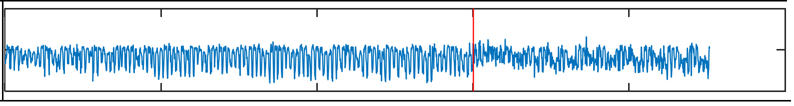
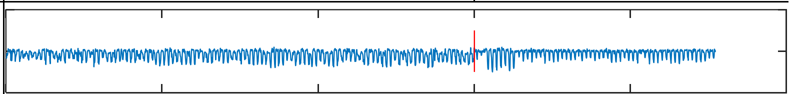
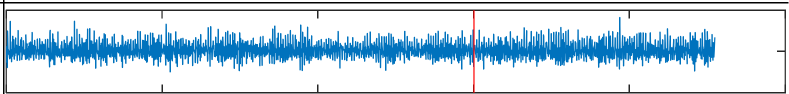
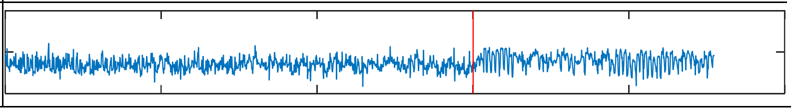
Sensor selection method	Sensor rank	Sensor number	Signal visualization (No-fault / Fault)
	4	2	
	5	1	
aVar-based PCA	1	3	
	2	2	
	3	1	
	4	4	
	5	5	
Aggregated index	1	3	
	2	4	
	3	2	
	4	1	
	5	5	

Table IV-9 Visualization of selected sensors in fault mode 3

Sensor selection method	Sensor rank	Sensor number	Signal visualization (No-fault / Fault)
PCA	1	4	
	2	3	
	3	5	

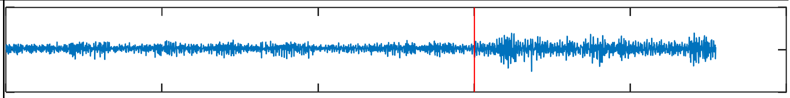
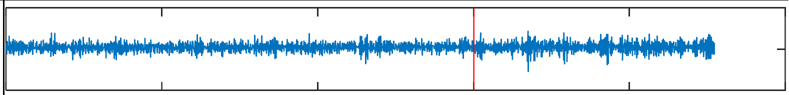
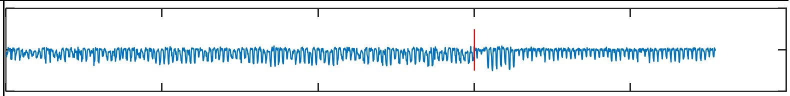
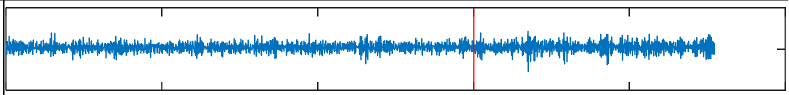
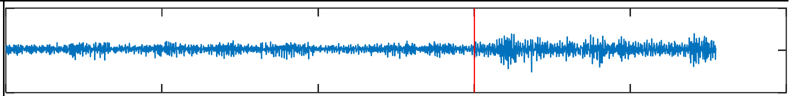
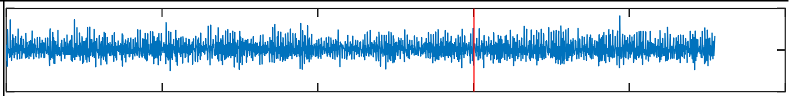
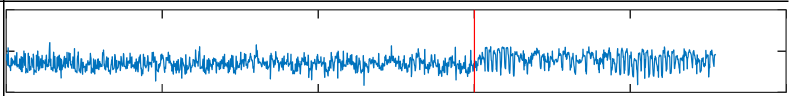
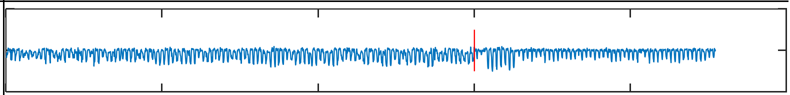
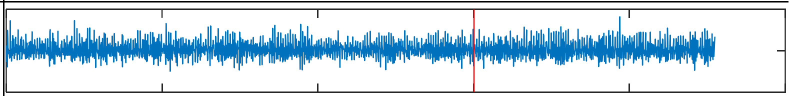
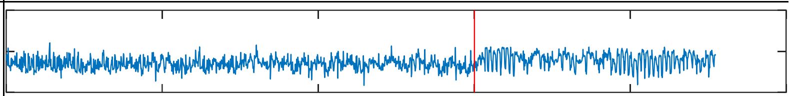
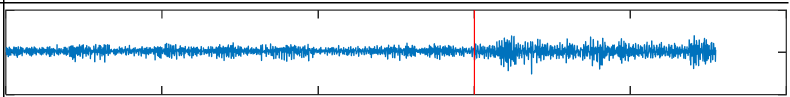
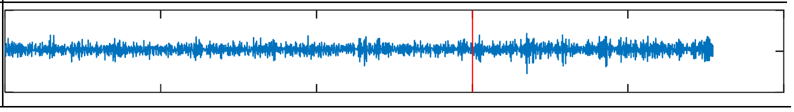
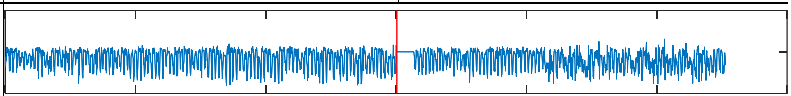
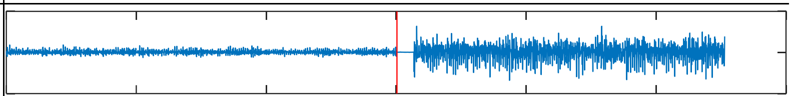
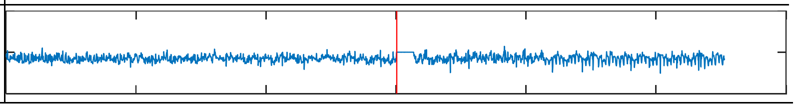
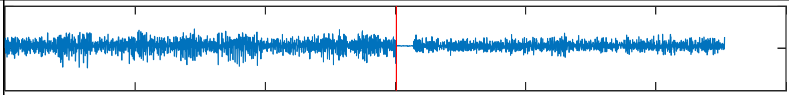
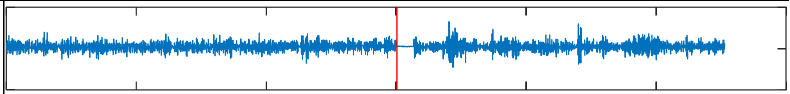
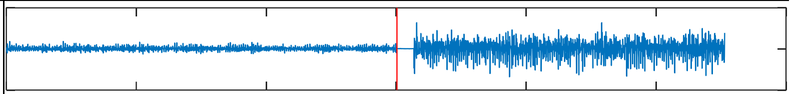
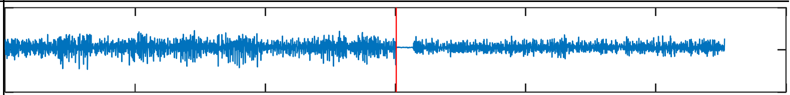
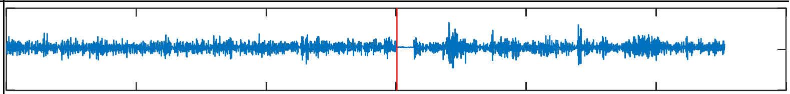
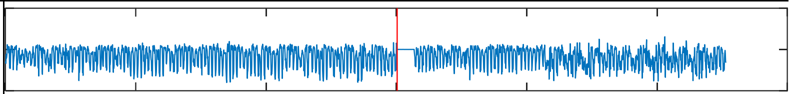
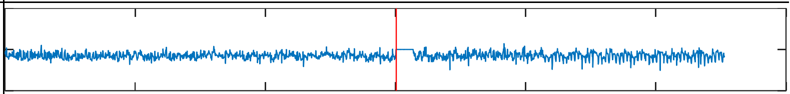
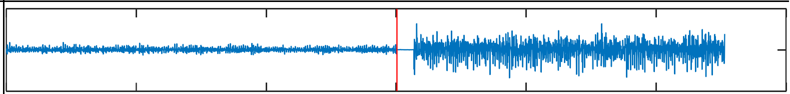
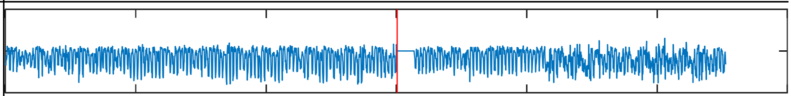
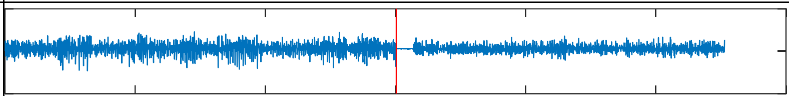
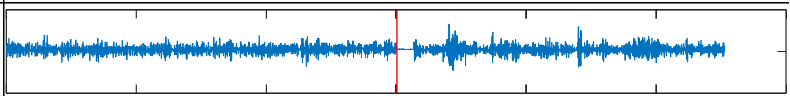
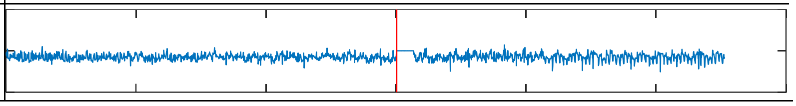
Sensor selection method	Sensor rank	Sensor number	Signal visualization (No-fault / Fault)
	4	2	
	5	1	
aVar-based PCA	1	**4**	
	2	**1**	
	3	**2**	
	4	**3**	
	5	**5**	
Aggregated index	1	4	
	2	3	
	3	5	
	4	2	
	5	1	

Table IV-10 Visualization of selected sensors in fault mode 4

Sensor selection method	Sensor rank	Sensor number	Signal visualization (No-fault / Fault)
PCA	1	4	
	2	3	
	3	5	

Sensor selection method	Sensor rank	Sensor number	Signal visualization	
			No-fault	Fault
	4	2		
	5	1		
aVar-based PCA	1	3		
	2	2		
	3	1		
	4	4		
	5	5		
Aggregated index	1	3		
	2	4		
	3	2		
	4	1		
	5	5		

First insight is that sensor ranks are same when using PCA in all fault modes: sensor4 (1^{st} rank), sensor3, sensor5, sensor2 and sensor1 (last rank). Sensors are ordered in ascending orders, which means the sensor on the top is the first rank. However, high-ranked sensors based on total variance does not show any characteristic signal distinguishing fault and no-fault state. Total variance of sensor is closely dependent on the location of failure points so the variance of sensor1, attached on the motor, is the lowest in all cases. Thus, when using PCA-based sensor selection sensor1 cannot be selected even with characteristic signal change.

Second, when using *aVar*-based PCA sensor selection, selected sensor order is totally different from the order by conventional PCA sensor selection. However, due to intrinsic characteristic of vibration data, sensor signals show repeated up-down trends even though total variance is around one. In this case, results show that signals of high-ranked sensor do not give any distinguishable trend or change in signal for fault and no-fault classification compared to low-ranked sensor.

Also, as listed in the Table IV-6, sparse impulse scores of every signal are similar, lower than 0.3 or around 0.3 score, which means there is no sparse impulse signal which is meaningful for classification. In this experiment, it is shown that using only *DI* score will be effective for classification than using *aVar*, PCA, or aggregated index. However, sensor selection by *DI* does not always give effective results for classification, like sensor3 in fault mode 4. In fault mode 2 and 3, sensor4 and sensor5 have lowest *DI* score, which means it is the most discernible sensor but those sensors do not show any clear difference between fault state and no-fault state referring the results of each signal.

As a result, two important insights were got from the results.

[1] Reducing number of sensors among small sensor sets might be meaningless or even it might delete any important data

[2] Needs for analyzing *aVar, DI, SI* in frequency domain not only in time domain

In this experiment, the number of sensors used in machine monitoring are not many compared to first case study, vehicle fault generator. Thus, from the beginning reducing number of sensors among small sensor sets might be meaningless or even it might delete any important data. Also, due to the intrinsic characteristics of sensor signals such as repeated vibration, further study considering frequency domain is needed.

This Page Intentionally Left Blank

V. Conclusion and Future research

Researches regarding fault diagnosis and prognosis have been studied for many centuries. As for fault diagnosis, classification performance plays an important role. However, existing sensor selection methods, which are dimension reduction techniques, has several limitations. First, conventional dimension reduction techniques using space transformation, it does not choose original sensors, instead it generates new axes axis so that it is hard to interpret the meaning of them. Also, the effectiveness and usefulness of using new axes in fault diagnosis should be considered. Second, in the case of dimension reduction by variable subset selection, original sensors are selected as a result. However, it is a kind of simple version of classification before conducting fault and no-fault classification. Randomly subset selection method generated sensor subsets randomly and adjust them to classification for finding best sensor subset so that it is computationally expensive. Therefore, simpler sensor selection method using original sensors is needed.

In this thesis, key sensor characteristics of signal are discussed and then developed a sensor selection methods based on the discussed indices considering (i) needs for dimension reduction without loss of original variable information, (ii) way to improve classification performance.

In summary, key contribution in this research is simplified. I proposed new index *SI* to emphasize the importance of the signals which *aVar* and *DI* cannot cover and were low-ranked in PCA sensor selection method but effective to fault and no-fault classification. As a result, classification performance increased. Hit rate was similar with PCA sensor selection and false alarm decreased a lot in proposed sensor selection aggregating PCA, *aVar*, *DI*, and *SI*.

There still exists challenging issues on the proposed sensor selection methods. The direction of future research can be summarized as follows. First, further analysis using proposed indices in frequency domain is needed. Second, analyzing the mathematical relationships between *DI* and *SI* are needed.

References

1. Al-Bugharbee, H., & Trendafilova, I. (2016). A fault diagnosis methodology for rolling element bearings based on advanced signal pretreatment and autoregressive modelling. *Journal of Sound and Vibration, 369*, 246-265.
2. Baek, S., Baek, W., Oh, H., & Kim, D. Y. (2016). *센서정보를 활용한 설비 및 제품 품질 예지보전 기술 동향*. Paper presented at the Proceedings of the CDE Conference, Pukyong National University in Republic of Korea.
3. Baek, S., & Kim, D. Y. (2017). Empirical Sensitivity Analysis of Discretization Parameters for Fault Pattern Extraction From Multivariate Time Series Data. *IEEE transactions on cybernetics, 47*(5), 1198-1209.
4. Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of chemometrics, 17*(3), 166-173.
5. Boashash, B. (2015). *Time-frequency signal analysis and processing: a comprehensive reference*: Academic Press.
6. Donoho, D. L., & Gavish, M. (2013). The optimal hard threshold for singular values is 4/√3. *arXiv preprint arXiv: 1305.5870*.
7. Ebrahimzadeh, A., & Ghazalian, R. (2011). Blind digital modulation classification in software radio using the optimized classifier and feature subset selection. *Engineering Applications of Artificial Intelligence, 24*(1), 50-59.
8. Gu, Q., Li, Z., & Han, J. (2012). Generalized fisher score for feature selection. *arXiv preprint arXiv:1202.3725*.
9. Hartmann, W. M. (2004). *Dimension reduction vs. variable selection.* Paper presented at the International Workshop on Applied Parallel Computing.
10. He, X., Cai, D., & Niyogi, P. (2006). *Laplacian score for feature selection.* Paper presented at the Advances in neural information processing systems.
11. Hossen, A., Al-Wadahi, F., & Jervase, J. A. (2007). Classification of modulation signals using statistical signal characterization and artificial neural networks. *Engineering Applications of Artificial Intelligence, 20*(4), 463-472.
12. Hu, W., Guo, M.-M., & Shi, L. (2012). *Research on fault tendency prediction of complex equipment based on multi-source information fusion.* Paper presented at the Quality, Reliability, Risk, Maintenance, and Safety Engineering (ICQR2MSE), 2012 International Conference on.
13. Izenman, A. J. (2013). Linear discriminant analysis *Modern multivariate statistical techniques* (pp. 237-280): Springer.
14. Janecek, A., Gansterer, W., Demel, M., & Ecker, G. (2008). *On the relationship between feature*

selection and classification accuracy. Paper presented at the New Challenges for Feature Selection in Data Mining and Knowledge Discovery.

15. Jimenez, L. O., & Landgrebe, D. A. (1998). Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 28*(1), 39-54.
16. Kano, M., & Nakagawa, Y. (2008). Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry. *Computers & Chemical Engineering, 32*(1), 12-24.
17. Kenne, J., & Boukas, E. (1997). *Production and corrective maintenance planning problem of a failure prone manufacturing system.* Paper presented at the American Control Conference, 1997. Proceedings of the 1997.
18. Kira, K., & Rendell, L. A. (1992). *The feature selection problem: Traditional methods and a new algorithm.* Paper presented at the Aaai.
19. Konar, P., & Chattopadhyay, P. (2011). Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs). *Applied Soft Computing, 11*(6), 4203-4211.
20. Kumar, V., & Minz, S. (2014). Feature Selection. *SmartCR, 4*(3), 211-229.
21. Liu, H., Liu, C., & Huang, Y. (2011). Adaptive feature extraction using sparse coding for machinery fault diagnosis. *Mechanical Systems and Signal Processing, 25*(2), 558-574.
22. Lopatka, J., & Pedzisz, M. (2000). *Automatic modulation classification using statistical moments and a fuzzy classifier.* Paper presented at the Signal Processing Proceedings, 2000. WCCC-ICSP 2000. 5th International Conference on.
23. Lu, B., Durocher, D. B., & Stemper, P. (2009). Predictive maintenance techniques. *IEEE Industry Applications Magazine, 15*(6).
24. Malik, M. A. K. (1979). Reliable preventive maintenance scheduling. *AIIE transactions, 11*(3), 221-228.
25. McInerny, S. A., & Dai, Y. (2003). Basic vibration signal processing for bearing fault detection. *IEEE Transactions on education, 46*(1), 149-156.
26. Minka, T. P. (2001). *Automatic choice of dimensionality for PCA.* Paper presented at the Advances in neural information processing systems.
27. Mörtsell, M., & Gulliksson, M. (2001). *An overview of some non-linear techniques in chemometrics*: Mitthögskolan, FSCN.
28. Nai, W., Zhang, F., Yu, Y., & Dong, D. (2015). A Prognosis Method Evaluating Fault Tendency Percentage for Key Equipment in Urban Rail Transit Signal System Based on Enhanced Naive Bayesian Classifying *CICTP 2015* (pp. 1623-1634).
29. Pal, M., & Foody, G. M. (2010). Feature selection for classification of hyperspectral data by SVM. *IEEE Transactions on Geoscience and Remote Sensing, 48*(5), 2297-2307.

30. Peter, W. T., Yang, W.-x., & Tam, H. (2004). Machine fault diagnosis through an effective exact wavelet analysis. *Journal of Sound and Vibration, 277*(4), 1005-1024.
31. Pollack, I., & Norman, D. A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic science, 1*(1-12), 125-126.
32. Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. *bioinformatics, 23*(19), 2507-2517.
33. Sagha, H., Millán, J. d. R., & Chavarriaga, R. (2011). *Detecting anomalies to improve classification performance in opportunistic sensor networks.* Paper presented at the Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on.
34. Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers, 31*(1), 137-149.
35. Swami, A., & Sadler, B. M. (2000). Hierarchical digital modulation classification using cumulants. *IEEE Transactions on communications, 48*(3), 416-429.
36. Tang, J., Alelyani, S., & Liu, H. (2014). Feature selection for classification: A review. *Data Classification: Algorithms and Applications*, 37.
37. Thompson, B. (2005). Canonical correlation analysis. *Encyclopedia of statistics in behavioral science*.
38. Trendafilov, N. T., & Jolliffe, I. T. (2007). DALASS: Variable selection in discriminant analysis via the LASSO. *Computational Statistics & Data Analysis, 51*(8), 3718-3736.
39. Wang, F., Zhang, Y., Zhang, B., & Su, W. (2011). *Application of wavelet packet sample entropy in the forecast of rolling element bearing fault trend.* Paper presented at the Multimedia and Signal Processing (CMSP), 2011 International Conference on.
40. Wang, Y., & Ma, X. (2012). *Optimal sensor selection for wind turbine condition monitoring using multivariate principal component analysis approach.* Paper presented at the Automation and Computing (ICAC), 2012 18th International Conference on.
41. Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems, 2*(1-3), 37-52.
42. Wolf, L., & Bileschi, S. (2005). *Combining variable selection with dimensionality reduction.* Paper presented at the Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on.
43. Wowk, V. (1991). *Machinery Vibration: Measurement and Analysis*: McGraw Hill.
44. Wu, S., WANG, C., HU, L., & HU, X.-y. (2006). Characteristic signal analysis and its identification of typical weld penetration status in laser welding. *TRANSACTIONS-CHINA WELDING INSTITUTION, 27*(7), 69.
45. Xiao, L., Li, J., N, A.-K. K., S, H. A., W, C. D., & A, E. E. (2013). Condition-based maintenance for continuously monitored degrading systems with multiple failure modes. *IIE Transactions, 45*(4),

422-435.

46. Xing, E. P., Jordan, M. I., & Karp, R. M. (2001). *Feature selection for high-dimensional genomic microarray data.* Paper presented at the ICML.
47. Ying, P., Ming, D., & Ming Jian, Z. (2010). Current status of machine prognostics in condition-based maintenance: a review. *The International Journal of Advanced Manufacturing Technology, 50*(1), 297-313.
48. Zhao, Z., Morstatter, F., Sharma, S., Alelyani, S., Anand, A., & Liu, H. (2010). Advancing feature selection research. *ASU feature selection repository*, 1-28.