Computers in Industry 62 (2011) 42–50

Contents lists available at ScienceDirect

Computers in Industry

journal homepage: www.elsevier.com/locate/compind

Using SVM based method for equipment fault detection in a thermal power plant

Kai-Ying Chen [a], Long-Sheng Chen [b,*], Mu-Chen Chen [c], Chia-Lung Lee [d]

[a] Department of Industrial Engineering and Management, National Taipei University of Technology, Taipei, Taiwan, ROC
[b] Department of Information Management, Chaoyang University of Technology, Taichung, Taiwan, ROC
[c] Institute of Traffic and Transportation, National Chiao Tung University, Taipei, Taiwan, ROC
[d] Shanghai FirstTech Company Limited, China

A R T I C L E I N F O

Article history:
Received 19 February 2009
Received in revised form 25 May 2010
Accepted 31 May 2010

Keywords:
Thermal power
Maintenance
Data mining
Support vector machines
Classification

A B S T R A C T

Due to the growing demand on electricity, how to improve the efficiency of equipment in a thermal power plant has become one of the critical issues. Reports indicate that efficiency and availability are heavily dependant upon high reliability and maintainability. Recently, the concept of e-maintenance has been introduced to reduce the cost of maintenance. In e-maintenance systems, the intelligent fault detection system plays a crucial role for identifying failures. Data mining techniques are at the core of such intelligent systems and can greatly influence their performance. Applying these techniques to fault detection makes it possible to shorten shutdown maintenance and thus increase the capacity utilization rates of equipment. Therefore, this work proposes a support vector machines (SVM) based model which integrates a dimension reduction scheme to analyze the failures of turbines in thermal power facilities. Finally, a real case from a thermal power plant is provided to evaluate the effectiveness of the proposed SVM based model. Experimental results show that SVM outperforms linear discriminant analysis (LDA) and back-propagation neural networks (BPN) in classification performance.

1. Introduction

Thermal power plants fired by fossil fuels are one of the primary sources of noxious greenhouse gas emissions, producing, carbon dioxide. Even so, they are still the major source of supplying electricity in Taiwan. According to the annual report of the Taiwan Power Company (TPC), the total power generation of their eight thermal power plants exceeds 70% of the total energy generated nowadays [3] in Taiwan. Consequently, due to growing demands on electricity, how to improve the efficiency of equipment in a thermal power plant has become a critical issue.

Huang et al. [6] indicated that the efficiency and availability depend heavily on high reliability and maintainability. In order to raise efficiency, the equipment of thermal power plants is becoming larger and more complex. However, due to lack of manpower and information resources, the diagnosis and repair of failed equipment cannot usually be performed immediately. From lots of published articles [58–62], we can find that to identify the failure types of steam turbines and their root causes is time consuming. It needs professional knowledge regarding materials and mechanical engineering. Generally speaking, thermal power plant engineers can merely handle routine or uncomplicated maintenance tasks. Additional tests and expert advice are additionally required from the technical support of original equipment manufacturers for complex fault diagnosis and maintenance, although these additional tests are often costly and involve some risk to equipment [8]. Hence it leads to long downtimes for equipment and causes significant production losses [11]. In order to reduce the cost of maintenance and risky experiments, the concept of e-maintenance has been introduced to identify the root cause of component failure, to reduce the failures of production systems, to eliminate costly unscheduled shutdown maintenances, and to improve productivity [12].

In an e-maintenance system, the intelligent fault detection system plays a crucial role for identifying failures. Data mining techniques are the core of such intelligent systems and can greatly enhance their performance [6,8,9]. Applying these techniques to fault detection makes it possible to eliminate additional tests or experiments which usually involve high expense and highly risk [8]. Recently, several data mining techniques such as artificial neural networks, fuzzy logic systems, genetic algorithms, and rough set theory have all been employed to assist the detection and condition monitoring tasks [4,10]. For example, Yang and Liu [8] presented a hybrid-intelligence data mining framework which involves an attribute reduction technique and rough set theory to diagnose the faults of boilers. Shu [45] established an interactive data mining approach based inference system to solve the basic technical challenge and speed up the discovery of knowledge in nuclear power plant. Besides, some related works designed data

* Corresponding author.
E-mail addresses: kychen@ntut.edu.tw (K.-Y. Chen), lschen@cyut.edu.tw (L.-S. Chen), ittchen@mail.nctu.edu.tw (M.-C. Chen).

0166-3615/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.compind.2010.05.013

mining based models for failure inspections, but not for fault predictions, such as the work of Yang and Liu [8]. Vast amounts of data describing process variables for boilers and turbines have been used for monitoring, control and over-limit alarms. Huang et al. [6] used principle component analysis (PCA) and T^2 statistics to inspect different types of faults in a thermal power plant. Prasad et al. [9] proposed a histogram based method to monitor and maximize the performance of thermal power plants. Therefore, building an intelligent system for the fault prediction of any thermal power plants most valuable equipment, namely turbines, has become necessary.

Since artificial intelligence techniques can improve prediction accuracy and decrease people involvement, it is very important to select an appropriate learning algorithm. In recent years, support vector machines (SVM) [15], has been considered as one of the most effective supervised learning algorithm in many pattern recognition problems [16–18]. It has been reported that SVM provides a better classification result than other methods such as neural networks or decision trees [1,14,19–21,23,35]. Moreover, SVM has been widely applied to fault detection and diagnosis in production environment. For examples, Hsu et al. [47] integrated a feature extraction technique, independent component analysis (ICA), into SVM to develop an intelligent fault detector for non-Gaussian in multivariate processes. Li et al. [48] combined another dimension reduction method, partial least squares (PLS), with SVM to increase the performance of on-line fault detection in batch processes. In the work of Zhang [49], both kernel independent component analysis (KICA, for non-Gaussian distribution) and kernel principal component analysis (KPCA, for Gaussian distribution) are used for fault detection in, named Tennessee Eastman process, which is a complex non-linear process created by Eastman Chemical Company. Mahadevan and Shah [50] utilized one-class SVM for fault detection and diagnosis and claimed that their approach outperformed principal components analysis (PCA) and dynamic principal components analysis (DPCA). From these works, we can know that SVM is one of effective fault detection approach in realistic industrial processes. In addition, SVM has been usually combined with feature extraction techniques including ICA, PLS, PCA, KPCA, KICA, and DPCA. But, by using feature extraction techniques, the transformed smaller feature space cannot be explainable and this is not good for searching root causes further.

Therefore, this work proposes a SVM based model which integrates a dimension reduction scheme and the SVM classifier is used to predict the failures of turbines. In this proposed model, in order to handle the huge amounts of collected data, correlation analysis (CA) and decision tree (DT) feature reduction methods have been introduced. Moreover, back-propagation neural network (BPN) and linear discriminant analysis (LDA) are utilized as the benchmarks for comparison purposes. Finally, a real case from a thermal power plant in Taiwan is provided to evaluate the effectiveness of the proposed SVM based model.

2. Thermal power plant

This section provides a brief introduction of a thermal power plant. In a power plant, the prime mover is steam driven. By heating, water is transformed into steam, and it is then condensed to push a turbine of power generators to produce electricity. Fig. 1 illustrates the basic steps in converting fossil fuels to electricity.

2.1. Equipment in a thermal power plant

The equipment of a thermal power plant is schematically shown in Fig. 2. The equipment can be classified into four major groups. As described below, they are the steam generator, the steam turbine generator, the electrical driven generator, and the monitoring and alarm system.

1. Steam generator: The steam generating boiler produces steam with high purity, pressure and temperature required for the steam turbine that drives the electrical generator. The generator includes boiler, water feeding system, fuel system, SCR, air heater, EP, FGD, etc.
2. Steam turbine generator: The steam turbine generator is used to transform the thermal energy to mechanical energy. The generator includes the turbine and the condensed system. It is the major piece of equipment of a thermal power plant.
3. Electrical driven generator: The electrical driven generator transforms the mechanical energy to the electrical energy. The generator includes electrical generator, exciter, transformer, etc.
4. Monitoring and alarm system: The system is used to monitor the above generators, and sounds the alarm if any abnormal event occurs.

Among them, the steam turbine generator is one of the most valuable pieces of equipment in a thermal power plant. Therefore, we focus on analyzing the failures of turbines in this work.

2.2. The important monitoring parameters and failure analysis of the steam turbine generator

The steam turbine generator is the most crucial piece of equipment in a thermal power plant. The turbine is a complex multi-axle system involving high-pressure generators, low-pressure generators, and exciter rotors [51]. The steam turbine blade is extremely complex since it must be flexible enough to change shape during operation in response to cold temperatures and the dynamic coupling effect [58]. Researchers have paid lots of attention on analyzing failures including failure types and causes [52,53,59,60]. Besides, Marco et al. [62] indicated that turbine startup is one of the critical problems in the operation of electrical power plants. Parka et al. [61] indicated that reducing environmental damage and increasing turbine efficiency are essential issues. Chen [51] indicated that reliable power generation and low maintenance costs are the major goals of power plant administration. He considered turbines are one of major parts for maintenance to enhance the efficiency of power plant equipment. Therefore, he presented an operational maintenance model by employing radio frequency identification technology. Moreover, in the work of Akturk and Gurel [52], the operational importance of turbines was heightened and realized. Therefore, to keep this generator operating smoothly without interruption by any faults, inspecting the data of parameters reported by the monitoring and alarm system is an important task.

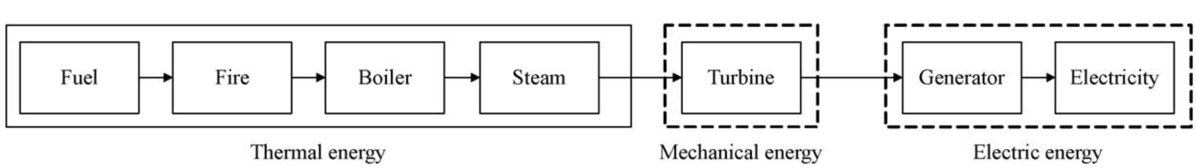

Fig. 1. Energy transformation of thermal power plant [31].

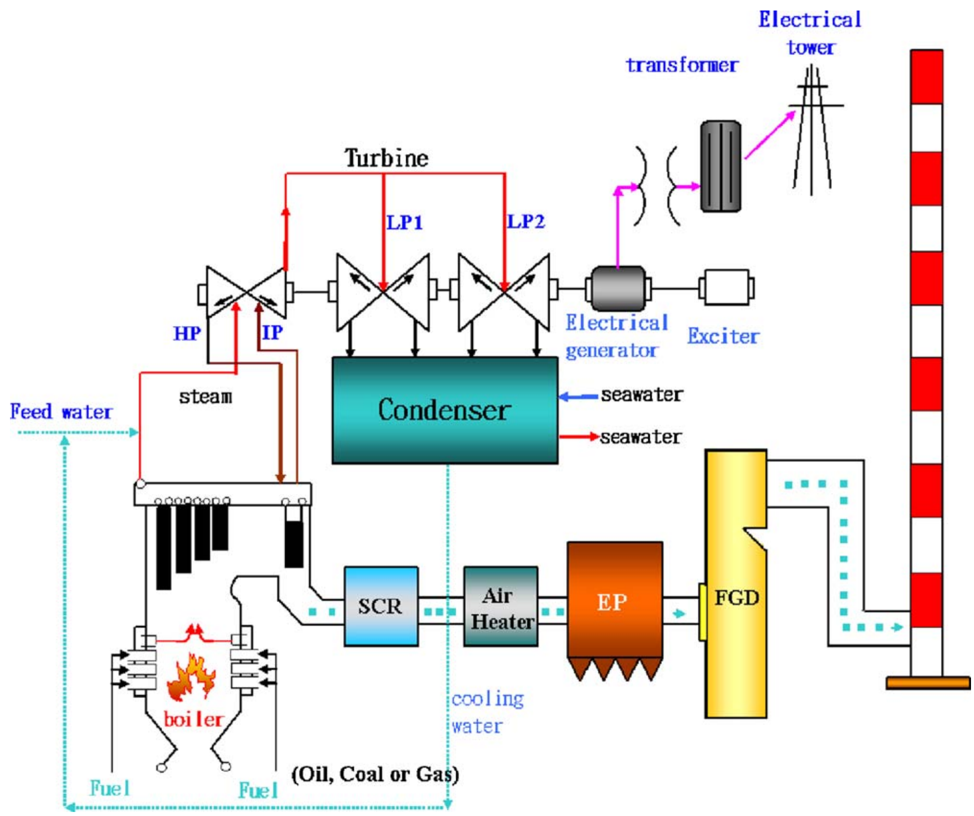

Fig. 2. A schematic illustration of thermal power plant [32].

The key monitoring parameters of the steam turbine generator include the temperature and pressure of the primary steam, the temperature and pressure of the reheated steam, vibration of the steam turbine generator and the rotation speed of turbine blades. Fig. 3 demonstrates that the temperature and pressure of primary steam and the rotation speed of the turbine shaft have a significant effect on the power load. Table 1 summarizes some of the causes with important monitoring parameters that lead to the failures of steam turbine generators. Engineers can maintain or repair the equipment with respect to the causes and failures.

Erosion of turbine blades may be caused by moisture [33]. If water gets into the steam and is blasted onto the blades, rapid impingement and erosion may occur that possibly leads to imbalance and catastrophic failure. The balance in large rotating steam turbines is vitally important to ensure the reliable operation of thermal power plants [34]. Most large steam turbines have sensors installed to measure the movement of the shafts in bearings. This condition monitoring can identify several potential problems and allow the repair of turbines to be planned before any problems become serious.

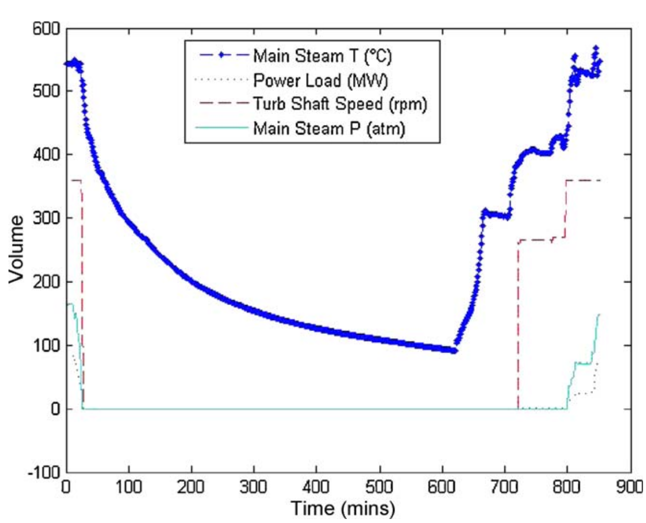

Fig. 3. The effect of parameters on the power load [32].

The generation of electricity requires precise speed control. Uncontrolled acceleration of turbine rotors may lead to an overspeed trip, which causes the blade valves controlling the flow of steam to the turbine to be closed. If so, the turbine may continue accelerating until it breaks apart, often spectacularly. The generator should rotate at constant synchronous speed according to the frequency of the electrical power system. The most common speed is 3600 revolutions per minute for a 60 Hz system [33].

The power company studied herein owns eight thermal power plants, and our case comes from a plant which owns four 500 MW oil/gas fired units. The experimental data are collected by the real-time monitor system.

Table 1
The causes of failures in steam turbine generator with monitoring parameters.

Monitoring parameters	Parameter variation	Potential failure
Pressure and temperature of primary steam and reheated steam	Abnormal increase	Failure in inlet steam of turbine
Pressure and temperature of primary steam and reheated steam	Abnormal decrease	Erosion of turbine blade
Vibration	Abnormal increase	Failure in bearing of turbine
Rotation speed	Over speed	Failure in blade of turbine

3. Implemental procedure of employed methodology

3.1. The proposed SVM based model

This section will introduce the detailed procedure of implementing our SVM based model. The employed model could be divided into five steps described as below:

Step 1: Identify monitoring parameters of turbines
Step 2: Data preprocessing
Step 2.1: Data cleaning (remove noisy or inconsistent data)
Step 2.2: Data transformation (normalize the data)
Step 3: Feature selection
Step 3.1: Implement correlation analysis
Step 3.2: Implement decision tree algorithm (C4.5)
Step 3.3: Acquire key parameters
Step 4: Construct SVM classifier
Step 3.1: Select kernel function
Step 3.2: Find optimal parameter settings
Step 3.3: Train SVM classifier
Step 5: Performance evaluation

Because modern power plants have been computerized, a huge amount of data could be automatically collected and stored. Therefore, the first step of our model is to identify the monitoring parameters which are related to turbine failure detection. The condition attributes (inputs) and the decision attribute (output) could be confirmed in step 1. Step 2 is data preparing phase. In this step, collected data should be prepared for implementing feature selection and constructing classifiers. We remove noisy and missing data, and then normalize these clean data. In step 3, two feature selection techniques including correlation analysis and decision tree have been utilized to reduce dimension of input data. By synthesizing both results of feature selection, the key condition attributes have been determined in this step. Next, the major task in step 4 is to build a SVM classifier including selecting kernel function, finding optimal parameter settings and training SVM. Finally, we will use a testing data to validate the effectiveness of the built SVM classifier. A more detailed discussion of our proposed approach is given in the following subsections.

3.2. Data preprocessing

After identifying input and output variables (step 1), data needs to be preprocessed. Step 2 is to clean data and transform data. Real-world data tend to incomplete, noisy, and inconsistent. In the step of data cleaning, this study attempts to remove the missing data, outliers, and correct inconsistencies in the data. Besides, different monitoring attribute has different scales. We need to normalize all attribute values into the same scale to avoid the influence of scales. All values of attributes are normalized to the interval [0,1] by using a min–max normalization equation, as expressed by Eq. (1). In this equation, $\max_i$ is the maximum and $\min_i$ is the minimum of the i-th attribute values, and v_{ij} is the value of i-th attribute of j-th object and v'_{ij} is the normalized value. In summary, data preprocessing techniques can improve the quality of the data, thereby helping to improve the accuracy and efficiency of data mining process.

$$v'_{ij} = \frac{v_{ij} - \min_i}{\max_i - \min_i} \quad (1)$$

3.3. Feature selection

Because the sizable data automatically collected in a thermal power is too huge to handle, dimension reduction techniques have been considered in this work. Typically, there are two kinds of algorithms to reduce the feature space in classification. The first one is feature selection which is to select a subset of most representative features from the original feature space. The second algorithm is feature extraction which is to transform the original feature space to a smaller one to reduce the dimension. Although feature extraction can reduce the dimension of feature space greatly compared with feature selection [26], the transformed smaller feature space cannot be explainable. Accordingly, feature selection algorithm is used in this study.

Dimension reduction via feature selection is one of the most fundamental steps in data processing [7,13]. Feature selection can be understood as choosing a subset of features that achieves the lowest error according to certain allowed losses [5]. A large feature set often contains redundant and irrelevant information, and can actually degrade the performance of classifier [44]. A number of soft computing approaches, such as neural networks, genetic algorithms (GA) [63], decision tree, rough sets [7,64], and correlation analysis have been widely used to remove irrelevant, unnecessary, and redundant attributes. However, when applying these feature selection to industry, we need to consider the computational cost and complexity. For the purpose of being easily used, we employ correlation analysis and decision tree to implement feature selection task in this study.

Correlation analysis is one of the common ways to select important features. This method is to evaluate the (linear) relationship among pair-wise inputs by using the correlation function [22]. Correlation analysis is a proven technique to remove redundant features, but it may fail when working with a low number of samples or when the assumed linear relationship does not exist. This method is employed as a base-line technique in order to explore the complexity of our particular feature selection problem. When using correlation analysis, first, we calculate the correlation coefficients between attributes, and then keep only one of the attributes which most highly correlates.

Another dimension reduction technique employed herein is the decision tree based feature selection method. A common understanding is that some learning algorithms have built-in feature selections such as decision trees [25]. When decision tree induction is used for feature selection, a tree is constructed from the given data. All attributes that do not appear in the tree are assumed to be irrelevant. The set of attributes appearing in the tree form the reduced subset of attributes [24]. In this study, we will synthesize the results of correlation analysis and decision tree to be our result of feature selection.

3.4. Learning algorithms

3.4.1. Support vector machines

In this work, SVM [15] has been employed as the learning algorithm due to its superior classification ability. SVM is a supervised learning technique and it can be used for classification and regression. The main advantages of SVM include the use of kernels (no need to acknowledge the non-linear mapping function), the absence of local minima (quadratic problem), the sparseness of solution and the generalization capability obtained by optimizing the margin [2].

Briefly speaking, SVM establishes a decision boundary between two classes by mapping the training data (through kernel functions) onto a higher dimensional space, and then finding the maximal margin hyperplane within that space. Finally, this hyperplane can be viewed as a classifier. The further introduction of SVM operations can be found in the following.

Giving n examples $S = \{x_i, y_i\}_{i=1}^{n}, \quad y_i \in \{-1, +1\}$, where x_i represents the condition attributes, y_i is the class label, and i is the number of examples. The decision hyperplane of SVM can be defined as (w, b), where w is a weight vector and b a bias. Let w_0 and b_0 denote the optimal values of the weight vector and bias.

Correspondingly, the optimal hyperplane can be written as

$$w_0^T x + b_0 = 0 \quad (2)$$

To find the optimum values of w and b, it is required to solve the following optimization problem.

$$\min_{w,b,\xi} \quad \frac{1}{2}w^T w + C\sum_{i=1}^{n}\xi_i$$
$$\text{Subject to} \quad y_i(w^T\phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (3)$$

where ξ is the slack variable, C is the user-specified penalty parameter of the error term ($C > 0$), and ϕ is the kernel function.

To sum up, SVM can change the original non-linear separation problem into a linear separation case by mapping input vector onto a higher feature space. On the feature space, the two-class separation problem is reduced to find the optimal hyperplane that linearly separates the two classes transformed into a quadratic optimization problem. In addition, several popular kernel functions are listed as Eqs. (4)–(7).

$$\text{Linear kernel} \quad K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i\mathbf{x}_j \quad (4)$$

Polynomial kernel of degreeg

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma\mathbf{x}_i\mathbf{x}_j + r)^g, \quad \gamma > 0 \quad (5)$$

Radial basis function

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\{-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2\}, \quad \gamma > 0 \quad (6)$$

$$\text{Sigmoid kernel} \quad K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma\mathbf{x}_i\mathbf{x}_j + r), \quad \gamma > 0 \quad (7)$$

Here, r, γ and g are kernel parameters and are user-defined. According the work of Hsu et al. [36], RBF kernel function is selected in this study. Readers can find more detailed information about SVM in Refs. [1,16,18,20].

In addition, the result of SVM is sensitive and might easily be influenced by parameter settings and the choice of kernel functions. To make sure of the optimal classification performance of an intelligent fault prediction system, BPN and LDA, which have shown their good classification ability in lots of published works, are employed as benchmarks.

3.4.2. Back-propagation neural networks

Neural networks implemented by using back-propagation learning algorithms have been widely applied in pattern recognition, function approximation, and optimization. In general, neural networks can be classified into two groups, feed-forward and feedback networks. In this work, we use a feed-forward network because of its superior classification ability. Besides, the back-propagation learning algorithm [28] is the best known training algorithm for neural networks and still one of the most useful. This iterative gradient algorithm is designed to minimize the mean square error (MSE) between the actual output of a multilayer feed-forward perceptron and the desired output. According to the rule of thumb and reports of available published papers, the number of hidden layers should be one or two. The back-propagation algorithm contains a forward pass and a backward pass. The forward pass attempts to obtain the activation value. The purpose of the backward pass is to adjust weights and biases according to the difference between the desired and actual network outputs. These two passes will go through iteratively until the network converges. The detailed information about BPN can be found in related references [29,30].

Table 2
The characteristics of dataset for thermal power plant.

No. of records	Input attributes	Target
10,822	29 continuous attributes	1 target attribute with 3 classes Normal: 67.75% Low: 24.65% Abnormal: 7.6%

3.4.3. Linear discriminant analysis

LDA projects high-dimensional data onto a low-dimensional space where the data can achieve maximum class separability [39]. The derived features in LDA are linear combinations of the original variables, where the coefficients are from the transformation matrix. The optimal projection in classical LDA is obtained by simultaneously minimizing the within-class distance while maximizing the between-class distance, thus achieving maximum class discrimination [40]. LDA has been successfully applied in many applications such as pattern recognition, image retrieval, face recognition, and so on [38,41–43].

In conventional statistical learning techniques, linear discriminant analysis [37] and logistic regression are very popular for pattern recognition [46]. But, LDA is stronger, involving ease of application once the initial model has been developed. In addition, LDA has no free parameters to be tuned and the extracted features are potentially interpretable under linearity assumptions [5]. As a result, LDA has also been adopted in this work.

4. Computational results

Table 2 provides a brief explanation about the data background, including data size, number of features, and class distribution. Totally, 10,822 examples are collected for further analysis. There are 29 input attributes (monitoring parameters) with continuous

Table 3
The summary of attributes.

Type	ID	Description	Notation	Type	ID	Description	Notation
Measure	M4471	BRG NO1 VIBRATION(X)	X1	Temperature	T4113	FIRST ST STM T PT 1	X16
Measure	M4499	TURBINE SHAFT SPEED	X2	Temperature	T4114	TURB COLD RHT T RS	X17
Pressure	P4111	TURB. PRIMARY IN P RS	X3	Temperature	T4115	TURB COLD RHT T LS	X18
Pressure	P4113	IMPULSE CHAMBER P	X4	Temperature	T4122	TURB HOT RHT INT RS	X19
Pressure	P4115	TURB. COLD RHT P LS	X5	Temperature	T4129	EXTR. 1 T AT TURB	X20
Pressure	P4120	RHTR. OUTLET P	X6	Temperature	T4132	EXTR. 3 T AT TURB	X21
Pressure	P4129	EXTR. 1 P AT TURB	X7	Temperature	T4144	CROSSOVER T TO LP TURB 1	X22
Pressure	P4132	EXTR. 3 P AT TURB3	X8	Temperature	T4151	HP GLAND STEAM T	X23
Pressure	P4144	CROSSVR P TO LP TURB 1	X9	Temperature	T4470	TURB OIL CLR OUT OIL T	X24
Pressure	P4145	CROSSVR P TO LP TURB 2	X10	Temperature	T4484	STM CHEST DP MTL T RH	X25
Pressure	P4151	HP GLAND STEAM P	X11	Temperature	T4485	STM CHEST DP MTL T LH	X26
Temperature	T4107	SEC SPHTR OUT T	X12	Temperature	T4486	STM CHEST SHLW MTL T RH	X27
Temperature	T4108	FIRST ST STM T PT 2	X13	Temperature	T4487	STM CHEST SHLW MTL T LH	X28
Temperature	T4109	MAIN STEAM T	X14	Temperature	T4488	HP FIRST STM T	X29
Temperature	T4111	TURB PRIMARY IN T RS	X15	Equipment status	F53	The equipment failure modes	Y

Table 4
Correlation analysis.

Input attributes		Correlation coefficient (R)
X4	X5	0.9990
X4	X7	0.9999
X4	X10	0.9989
X25	X26	0.9995
X25	X27	0.9998
X25	X28	0.9996
X16	X17	1.0000
X13	X20	0.9997
X12	X18	0.9978
X3	X6	0.9951
X19	X29	0.9903

values and 1 target attribute with three-class labels (normal, low and abnormal). For classification model building, 70% of the examples are used as a training set and 30% as a test set. Table 3 summarizes the attributes of the thermal power plant dataset.

4.1. Feature selection

The main purpose of feature selection is to remove irrelevant or redundant attributes and improve the performance of classification. This section provides the results of correlation analysis and decision tree based feature selection techniques. Table 4 shows some correlation coefficient between/among attributes. The classification may not be influenced if we merely keep one of those highly correlated attributes for further analysis. For example, X5, X7, and X10 (X26, X27, and X28) are highly correlated with X4 (X25). Therefore, X5, X7, and X10 (X26, X27, and X28) can be removed. As a result, 11 attributes including X5, X7, X10, X26, X27, X28, X17, X20, X18, X6, and X29 are considered as irrelevant features. As a result, 18 attributes are kept for building classifiers.

In addition, Table 5 lists the knowledge rules extracted by decision tree (C4.5). In the decision tree based feature selection technique, the attributes appearing in nodes of a constructed tree can be considered as important variables for classifying objects. According to the results shown in Table 5, except for 5 attributes (X3, X4, X21, X23, and X27), 24 attributes which do not appear in the tree can be removed.

In order to clearly demonstrate the results, we use the notation "Raw", "DT", and "CA" to represent the original data and the feature selection results of decision tree and correlation analysis, respectively. Moreover, the union of "DT" and "CA" is employed as another data set which is denoted as "DT + CA". Table 6 shows the detailed illustration of the employed datasets.

4.2. Building classifiers and performance evaluation

To perform SVM, RBF which is one of most widely used kernel functions in SVM applications [54–56] has been employed due to some desirable properties. For instance, RBF includes other kernels as special cases and avoids difficulties associated with very large numbers because its values range between 0 and 1. Moreover, it is well known, that the predictive performance of a SVM depends heavily upon an appropriate choice of its parameters settings [57]. Therefore, using RBF kernel function only needs to tune of its two parameters C and γ, which facilitate adapting the classifier to a particular task.

In this research, we use the LIBSVM version 2.82 [27], which is an integrated tool for support vector classification and regression, and is available at http://www.csie.ntu.edu.tw/~cjlin/libsvm. In LIBSVM, the optimal settings of these two parameters can be found by using grid search. Readers can find more detailed information regarding gird search in Ref. [36]. In BPN, the parameter settings and optimal structure of neural network are obtained by a trial-

Table 5
Extracted knowledge rules of decision tree.

Rule no.	IF	AND	THEN		Important variable
1	X3 < 70.96195	X4 < 16.33355	NODE: 4 No.: 416	Abnormal: 100.0% Low: 0.0% Normal: 0.0%	X3, X4
2	16.5336 <= X4	–	NODE: 7 No.: 3034	Abnormal: 0.0% Low: 0.0% Normal: 100.0%	X4
3	X4 < 16.2259	70.96195 <= X3	NODE: 8 No.: 1024	Abnormal: 0.0% Low: 99.7% Normal: 0.3%	X3, X4
4	X27 < 500.09	16.33355 <= X4 < 16.5336	NODE: 10 No.: 18	Abnormal: 0.0% Low: 100.0% Normal: 0.0%	X4, X27
5	500.09 <= X27	16.33355 <= X4 < 16.5336	NODE: 11 No.: 295	Abnormal: 0.0% Low: 7.5% Normal: 92.5%	X4, X27
6	X21 < 365.687	16.2259 <= X4 < 16.33355 70.96195 <= X3	NODE: 14 No.: 304	Abnormal: 0.0% Low: 12.2% Normal: 87.8%	X3, X4, X21
7	X23 < 316.973	365.687 <= X21 16.2259 <= X4 < 16.33355 70.96195 <= X3	NODE: 28 No.: 100	Abnormal: 0.0% Low: 99.0% Normal: 1.0%	X3, X4, X21, X23
8	316.973 <= X23	365.687 <= X21 16.2259 <= X4 < 16.33355 70.96195 <= X3	NODE: 29 No.: 151	Abnormal: 0.0% Low: 80.1% Normal: 19.9%	X3, X4, X21, X23

Table 6
Illustration of different employed data sets in experiments.

Data set	No. of attributes	Employed attributes	Illustration
Raw	29	X1-X29	Original data
DT	5	X3, X4, X21, X23, X27	Keep the attributes which appear in constructed decision trees
CA	18	X1, X2, X3, X4, X8, X9, X11, X12, X13, X14, X15, X16, X19, X21, X22, X23, X24, X25	After removing those who owns high correlation between each other
DT + CA	19	X1, X2, X3, X4, X8, X9, X11, X12, X13, X14, X15, X16, X19, X21, X22, X23, X24, X25, X27	The union of "Re" and "DT" data sets

Table 7
The parameters and structure setting in BPN.

Data set	Structure	Learning rate	Momentum	Iterations
Raw	29-15-1	0.02	0.9	5000
DT	5-10-1	0.02	0.9	5000
CA	18-15-1	0.02	0.9	5000
DT + CA	19-15-1	0.02	0.9	5000

Table 8
The summary of SVM, BPN and LDA.

Method		SVM		BPN		LDA	
Performance		Training (%)	Testing (%)	Training (%)	Testing (%)	Training (%)	Testing (%)
Dataset and attribute no.							
Raw	29	93.64	93.13	87.71	86.50	83.62	76.71
DT	5	90.94	90.20	82.87	81.62	81.20	74.91
CA	18	92.16	91.58	87.49	88.13	83.02	76.23
DT + CA	19	93.23	92.79	87.70	86.04	82.97	75.95
Average		92.49	91.93	86.44	85.57	82.70	75.95

and-error method whose results can be found in Table 7. Learning rate and momentum are set as 0.02 and 0.9, respectively. The number of training iterations is set to 5000.

Table 8 summarizes the results of SVM, BPN, and LDA on four different datasets. Without implementing feature selection techniques (raw dataset), SVM has the highest classification accuracy (93.13%) compared with BPN (86.50%) and LDA (76.71%). Generally speaking, from Fig. 4, it is easy to find that SVM is slightly better than BPN and greatly outperforms LDA in all four datasets. On average, SVM has a 91.93% classification accuracy, and it is better than those of BPN (85.57%) and LDA (79.95%). In fact, we have statistical evidence to support these results. Two hypothesis tests also have been provided as follows.

$$H_0: \mu_{\text{SVM}} \leq \mu_{\text{BPN}} \qquad H_1: \mu_{\text{SVM}} > \mu_{\text{BPN}} \tag{8}$$

As a result, the p-value is 0.015 (<0.05). Then we draw a conclusion of rejecting H_0. We have 95% confidence to believe that SVM is significantly better than BPN.

$$H_0: \mu_{\text{BPN}} \leq \mu_{\text{LDA}} \qquad H_1: \mu_{\text{BPN}} > \mu_{\text{LDA}} \tag{9}$$

As a result, the p-value is 0.03 (<0.05). We can therefore reject the null hypothesis (H_0). We have 95% confidence to believe that BPN is better than LDA. Finally, we can conclude that SVM has a better

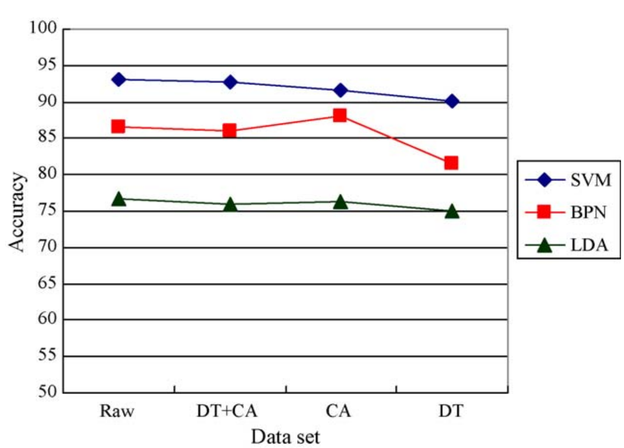

Fig. 4. The comparisons of SVM, BPN, and LDA.

performance than BPN and LDA in this case. Therefore, SVM is suitable to be selected as a classifier for analyzing turbine failures.

Regarding the performance of feature selection, we found that the performance will drop slightly with the reduction of attributes. DT technique can remarkably reduce the dimensionality size from 29 to 5. CA and DT + CA techniques also can do this job from 29 to 18 and to 19, respectively. But, DT technique has the largest drop in classification performances (SVM: ↓2.93%; BPN: ↓4.84%; LDA: ↓1.8%) among three feature selection methods. Considering both performance and dimension reduction, the results indicated that DT + CA not only can reduce the amount of monitoring parameters (from 29 to 19), but also can keep the classification performance (SVM: ↓0.34%; BPN: ↓0.46%; LDA: ↓ 0.76%). Although the accuracy slightly decreases, the reduction of input variable reaches up to 82.76%. A little loss of accuracy can shorten training procedure and save data storage space.

5. Conclusions

This work proposed a SVM based model for predicting failures of turbines in a thermal power plant. In order to handle the huge amount of collected data, the proposed SVM based model integrates feature selection techniques. Finally, a real-world data from a thermal power company has been employed to evaluate the effectiveness of our proposed model. By comparing performances and effectiveness of SVM, BPN, and LDA for dealing with inspection data of a thermal power plant, the experimental results indicated the performance of the classifier (SVM) of our model as being superior to those of BPN and LDA. The SVM based model can successfully detect the types of turbine faults with a high degree of accuracy (greater than 90%). Our proposed method can assist online engineers to find failure types without the support of original equipment manufacturers, which are often expensive and time consuming. It can dramatically shorten the shutdown maintenance time and thus increase the capacity utilization rate of turbines.

In feature selection, although these dimension reduction approaches will result in a slight loss of classification accuracy, they are within acceptable limits, compared with the advantage of saving the data processing time and computation time of learning algorithms in the training phase. In this work, we proposed a simple and clear feature selection technique which could be easily followed by readers in industry. However, if companies do not considering computational costs, Support vector machine recursive feature extraction (SVM-RFE) algorithm and genetic algorithms (GA) might result in better performances. It could be a possible direction of further studies. Moreover, although our method can predict the fault types, further analysis of failures is necessary to discover root causes. That can be a potential direction for future works. In addition, the proposed model needs suitable equipments such as sensors and data storage devices to provide data for training. This is the limitation of our model.

Acknowledgment

The authors would like to thank the National Science Council of Taiwan, ROC for supporting this research in part under Contract No. NSC 95-2218-E-027-001.

REFERENCES

[1] B.M. Asl, S.K. Setarehdan, M. Mohebbi, Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal, Artificial Intelligence in Medicine (2008), doi:10.1016/j.artmed.2008.04.007.
[2] A.S. Cerqueira, D.D. Ferreira, M.V. Ribeiro, C.A. Duque, Power quality events recognition using a SVM-based method, Electric Power Systems Research 78 (9) (2008) 1546–1552.

[3] C.-F. Chien, W.-C. Chen, F.-Y. Lo, Y.-C. Lin, A case study to evaluate the productivity changes of the thermal power plants of the Taiwan power company, IEEE Transactions on Energy Conversion 22 (3) (2007) 680–688.
[4] A. Evsukoff, S. Gentil, Recurrent neuro-fuzzy system for fault detection and isolation in nuclear reactors, Advanced Engineering Informatics 19 (1) (2005) 55–66.
[5] J. Gomez-Sanchis, L. Gomez-Chova, N. Aleixos, G. Camps-Valls, C. Montesinos-Herrero, Hyperspectral system for early detection of rottenness caused by Penicillium digitatum in mandarins, Journal of Food Engineering 89 (1) (2008) 80–86.
[6] X. Huang, H. Qi, X. Liu, Implementation of fault detection and diagnosis system for control systems in thermal power plants, in: Proceedings of the 6th World Congress on Intelligent Control and Automation, Dalian, China, June 21–23, 2006.
[7] R.W. Swiniarski, L. Hargis, Rough sets as a front end of neural-networks texture classifiers, Neurocomputing 36 (2001) 85–102.
[8] P. Yang, S.S. Liu, Fault Diagnosis for boilers in thermal power plant by data mining, in: Proceedings of Eighth International Conference on Control, Automation, Robotics and Vision, Kunming, China, December 6–9, 2004.
[9] G. Prasad, E. Swidenbank, W. Hogg, A novel performance monitoring strategy for economical thermal power plant, IEEE Transactions on Energy Conversion 14 (3) (1999) 802–809.
[10] M.C. Garcia, M.A. Sanz-Bobi, J. Pico, SIMAP: intelligent system for predictive maintenance application to the health condition monitoring of a windturbine gearbox, Computers in Industry 57 (2006) 552–568.
[11] T. Han, B.-S. Yang, Development of an e-maintenance system integrating advanced techniques, Computers in Industry 57 (2006) 569–580.
[12] B. Iung, A.C. Marquez, Editorial: special issue on e-maintenance, Computers in Industry 57 (2006) 473–475.
[13] I.K. Fodor, C. Kamath, Dimension reduction techniques and classification of bent double galaxies, Computational Statistics and Data Analysis 41 (1) (2002) 91–122.
[14] M.-D. Yang, T.-C. Su, Automated diagnosis of sewer pipe defects based on machine learning approaches, Expert Systems with Applications 35 (3) (2008) 1327–1337.
[15] V. Vapnik, Statistical Learning Theory, Springer, Berlin, 1998.
[16] Y.H. Liu, H.P. Huang, C.H. Weng, Recognition of electromyographic signals using cascaded kernel learning machine, IEEE/ASME Transactions on Mechatronics 12 (3) (2007) 253–264.
[17] K. Polat, S. Günes, A. Arslan, A cascade learning system for classification of diabetes disease: generalized discriminant analysis and least square support vector machine, Expert Systems with Applications 34 (1) (2008) 482–487.
[18] Y.H. Liu, Y.T. Chen, Face recognition using total margin-based adaptive fuzzy support vector machines, IEEE Trans Neural Networks 18 (1) (2007) 178–192.
[19] S. Theodoridis, K. Koutroumbas, Pattern Recognition, 2nd ed., Elsevier Academic Press, San Diego, 2003.
[20] J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery 2 (1998) 121–167.
[21] L.-S. Chen, C.-T. Su, Using granular computing model to induce scheduling knowledge in dynamic manufacturing environments, International Journal of Computer Integrated Manufacturing 21 (5) (2008) 569–583.
[22] J.F. Hair, R.E. Anderson, R.L. Tatham, W. Black, Multivariate Data Analysis, 4th ed., Prentice Hall, Englewood Cliffs, 1995.
[23] I. Hmeidi, B. Hawashin, E. El-Qawasmeh, Performance of KNN and SVM classifiers on full word Arabic articles, Advanced Engineering Informatics 22 (1) (2008) 106–111.
[24] J. Han, M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, San Francisco, 2001.
[25] J. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann, Bari, 1993.
[26] T. Liu, Z. Chen, B. Zhang, W. Ma, G. Wu, Improving text classification using local latent semantic indexing, Proceedings of the Fourth IEEE International Conference on Data Mining (2004) 162–169.
[27] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, 2001. Software available from: http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[28] D.E. Rumelhart, J.L. McClelland, Parallel Distributed Processing, MIT Press and the PDP Research Group, Cambridge, 1986.
[29] D.W. Philip, Neural Computing: Theory and Practice, Van Nostrand Reinhold, New York, 1992.
[30] A.J. Freeman, M.D. Skapura, Neural Networks: Algorithms, Applications, and Programming Techniques, Addison Wesley, MA, 1992.
[31] http://www.centreforenergy.com/.
[32] http://www.taipower.com.tw/.
[33] http://en.wikipedia.org/wiki/Steam_turbine.
[34] http://www.energymanagertraining.com/power_plants/steam_turbine.htm.
[35] R.B. Chinnam, Support vector machines for recognizing shifts in correlated and other manufacturing processes, International Journal of Production Research 40 (17) (2002) 4449–4466.
[36] C.-W. Hsu, C.-C. Chang, C.-J. Lin, A practical guide to support vector classification, 2006. http://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html.
[37] R.A. Fisher, The use of multiple measurements in taxonomic problems, Annals of Eugenics 7 (1936) 179–188.
[38] A. Martinez, A. Kak, PCA versus LDA, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2) (2001) 228–233.
[39] R. Duda, P. Hart, D. Stork, Pattern Classification, Wiley, 2000.
[40] Y. Ye, Least squares linear discriminant analysis, Proceedings of the 24th International Conference on Machine Learning (2007) 1087–1093.
[41] D. Swets, J. Weng, Using discriminant eigenfeatures for image retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (8) (1996) 831–836.
[42] P. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs fisherfaces: recognition using class specific linear projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.
[43] S. Dudoit, J. Fridlyand, T.P. Speed, Comparison of discrimination methods for the classification of tumors using gene expression data, Journal of the American Statistical Association 97 (2002) 77–87.
[44] O. Oyeleye, E.A. Lehtihet, A classification algorithm and optimal feature selection methodology for automated solder joint defect inspection, Journal of Manufacturing Systems 17 (1998) 251–262.
[45] Y. Shu, Inference of power plant quake-proof information based on interactive data mining approach, Advanced Engineering Informatics 21 (3) (2007) 257–267.
[46] V.S. Desai, J.N. Crook, G.A. Overstreet Jr., A comparison of neural networks and linear scoring models in the credit union environment, European Journal of Operational Research 95 (1996) 24–37.
[47] C.-C. Hsu, M.-C. Chen, L.-S. Chen, Intelligent ICA–SVM fault detector for non-Gaussian multivariate process monitoring, Expert Systems with Applications 37 (4) (2010) 3264–3273.
[48] Y. Li, Z. Wang, J. Yuan, On-line fault detection using SVM-based dynamic MPLS for batch processes, Chinese Journal of Chemical Engineering 14 (6) (2006) 754–758.
[49] Y. Zhang, Enhanced statistical analysis of nonlinear processes using KPCA, KICA and SVM, Chemical Engineering Science 64 (5) (2009) 801–811.
[50] S. Mahadevan, S.L. Shah, Fault detection and diagnosis in process data using one-class support vector machines, Journal of Process Control 19 (10) (2009) 1627–1639.
[51] T.-L. Chen, Real-time turbine maintenance system, Expert Systems with Applications 36 (4) (2009) 8676–8681.
[52] M.S. Akturk, S. Gurel, Machining conditions-based preventive maintenance, International Journal of Production Research 45 (8) (2007) 1725–1743.
[53] N. Vardar, A. Ekerim, Failure analysis of gas turbine blades in a thermal power plant, Engineering Failure Analysis 14 (4) (2007) 743–749.
[54] K. Coussement, K.D. Poel, Churn prediction in subscription services: an application of support vector machines while comparing two parameter selection techniques, Expert Systems with Applications 34 (1) (2008) 313–327.
[55] S.S. Keerthi, C.-J. Lin, Asymptotic behaviors of support vector machines with Gaussian kernel, Neural Computation 15 (7) (2003) 1667–1689.
[56] S. Lessmann, S. Voß, A reference model for customer-centric data mining with support vector machines, European Journal of Operational Research 199 (2) (2009) 520–530.
[57] C.-T. Su, C.-H. Yang, K.-H. Hsu, W.-K. Chiu, Data mining for the diagnosis of type II diabetes from three-dimensional body surface anthropometrical scanning data, Computers & Mathematics with Applications 51 (2006) 1075–1092.
[58] A. Pesiridis, R.F. Martinez-Botas, Experimental evaluation of active flow control mixed-flow turbine for automotive turbocharger application, Transactions American Society of Mechanical Engineers Journal of Turbomachinery 129 (1) (2007) 44–52.
[59] G. Das, S.G. Chowdhury, A.K. Ray, S.K. Das, D.K. Bhattacharya, Turbine blade failure in a thermal power plant, Engineering Failure Analysis 10 (1) (2003) 85–91.
[60] J.-C. Chang, Y.-H. Yun, C. Choi, J.-C. Kim, Failure analysis of gas turbine buckets, Engineering Failure Analysis 10 (5) (2003) 559–567.
[61] D.H. Parka, G.M. Jungb, J.K. Yum, Cost minimization for periodic maintenance policy of a system subject to slow degradation, Reliability Engineering and System Safety 68 (2000) 105–112.
[62] A. Marco, A. Lopez, C.H. Flores, E.G. Garcia, An intelligent tutoring system for turbine startup training of electrical power plant operators, Expert Systems with Applications 24 (2003) 95–101.
[63] F. Zhu, S. Guan, Feature selection for modular GA-based classification, Applied Soft Computing 4 (2004) 381–393.
[64] R.W. Swiniarski, A. Skowron, Rough set methods in feature selection and recognition, Pattern Recognition Letters 24 (2003) 833–849.

Kai-Ying Chen received the B.S., M.S. and Ph.D. degrees in Mechanical Engineering from National Taiwan University in 1988, 1990 and 1996 respectively. After graduating, he served in Industrial Technology Research Institute as an engineer for 4 years. Dr. Chen is currently an associate professor with the Department of Industrial Engineering and Management in National Taipei University Technology, Taipei, TAIWAN. His research interests are in the field of shop floor control, manufacturing information system, data mining, petri-net modeling and RFID applications.

Long-Sheng Chen received his Ph.D. in Department of Industrial Engineering and Management, National Chiao Tung University, Taiwan in 2006, and his B.S. and M.S. degrees both in industrial management from National Cheng Kung University, Tainan, Taiwan in 1998 and 2000, respectively. He is currently an assistant professor with the Department of Information Management, Chaoyang University of Technology, Taiwan. His teaching and research interests include data mining, quality management, new service creation, class imbalance problems and neural networks applications.

Mu-Chen Chen received his Ph.D. and M.Sc. degrees both in Industrial Engineering and Management from National Chiao Tung University, and his B.S. degree in Industrial Engineering from Chung Yuan Christian University. He is currently a professor in the Institute of Traffic and Transportation at National Chiao Tung University, Taiwan. His teaching and research interests include data mining, logistics and supply chain management and meta-heuristics.

Chia-Lung Lee got his M.Sc. degree from Institute of Industrial Engineering and Management, National Taipei University of Technology, Taipei, Taiwan. His research interests include data mining and equipment failure analysis