Int J Adv Manuf Technol (2013) 68:763–773
DOI 10.1007/s00170-013-4797-0

ORIGINAL ARTICLE

Intelligent fault diagnosis and prognosis approach for rotating machinery integrating wavelet transform, principal component analysis, and artificial neural networks

Zhenyou Zhang · Yi Wang · Kesheng Wang

Received: 19 April 2012 / Accepted: 24 January 2013 / Published online: 21 February 2013

Abstract This paper proposes a new approach for rotating machinery which integrates wavelet transform (WT), principal component analysis (PCA), and artificial neural networks (ANN) to classify the fault and predict the conditions of components, equipment, and machines. The standard deviation of wavelet coefficients are extracted from processed historical signals of manufacturing equipment as features. Then, the features are analyzed by PCA and several new principal features obtained from original features can be used as inputs to train ANN. After training, the conditions and degradations of components and machines can be predicted, and the fault of them can be classified if it exists, by the trained ANN using the same kinds of principal features extracted from real time signals. A case study is used to evaluate the proposed method and the result indicates its higher efficiency and effectiveness comparing to traditional methods.

Keywords Fault diagnosis and prognosis · Wavelet packet coefficient · Artificial neural network · Principal component analysis

1 Introduction

The requirements for equipment are becoming more complex and complicated in the development of modern manufacturing technology. The trouble-free operation of equipment is very dependent upon the trouble-free operation of all its components [1]. During a system breakdown, up to 80 % of downtime is spent to locate the source of the fault [2]. Only a small fraction of the downtime is spent to maintain the components. This sparks a vast interest to study the corresponding intelligent diagnostic techniques and system for widespread applications [3]. Thus, the function of diagnosis is minimizing the resources, such as time, to spend on identifying the cause of the observed degradations [4]. Fault prognosis should be a vital part of the monitoring system [4]. In a real system, the prognosis is as important as diagnosis of fault in order to avoid the unplanned breakdown and reduce the cost of maintenance.

The crucial part of the system is detecting faults early enough to make optimal maintenance decision. The diagnosis and prognosis part of the system should be able to forecast the remaining useful life of equipment, the operating time between detection, and an unacceptable level of degradation. Currently, various sensors have been developed and adopted to monitor and record information such as temperature, pressure, sound, and vibration in manufacturing system [5]. In order to solve the practical maintenance problems of mechanical systems, many data mining methodologies, such as statistical inference, decision tree, rough sets, support vector machine (SVM), and various hybrid methods have been introduced and validated [6–10]. Artificial neural network (ANN) is selected as the core technology due to:

1. ANN deal with complex problems without sophisticated and specialized knowledge,
2. ANN is an effective classification technique,
3. ANN can deal with nonlinear systems and low operational response times after the learning phase.

Many methods based on ANN have been developed for online monitoring with knowledge discovery, novelty detection, and learning abilities [11–13]. Other technologies such

Z. Zhang · K. Wang (✉)
Department of Production and Quality Engineering, Norwegian University of Science and Technology, Trondheim, Norway
e-mail: kesheng.wang@ntnu.no

Z. Zhang
e-mail: zhenyou.zhang@ntnu.no

Y. Wang
University of Sunderland, Sunderland, UK
e-mail: yi.wang@sou.ac.uk

as fuzzy logic system, genetic algorithms, and hybrid computational intelligence systems were applied in fault diagnosis for a centrifugal pump, respectively [14]. Lee et al. [4] developed an intelligent prognostics and e-maintenance system named “Watchdog Agent” with the method of statistical matching and performance signature and SVM-based diagnostic tool. There are also many signal processing techniques for fault diagnosis and prognosis. Parameters in time domain [15–18], frequency domain [19–21], and wavelet domain [22–24] are extracted as features to train ANN to classify the fault of monitored components. Because a large number of sensors are used in real system, the features may become extremely large. Principal component analysis (PCA) was used to reduce the dimension of the features without decrease the accuracy [25, 26].

However, there are very few literatures on integrating wavelet domain, PCA, and ANN in fault diagnosis and prognosis area [23] and also very few literature can predict degradation of monitored components. The major contribution of this paper is to combine the three techniques. Accurate condition for each component can be obtained from the real time signals and this condition could be updated online. Based on this online condition of equipment, the proposed method can provide an intelligent decision making tool to analyze the data from sensors enabling working condition prediction which has the ability to monitor, diagnose, and predict the condition of manufacturing systems and its processes. It can make maintenance decision based on current information to prevent occurrence and development of failure effectively, ensure the safety of equipment and personnel, and reduce economic lost caused by failure. It is hoped that it can help companies to reach near-zero breakdown performance with low cost of maintenance and loss of product, and improving reliability, safety, and productivity.

Based on the framework of the Intelligent Fault Diagnosis and Prognosis Systems, a case for diagnosis and prognosis of unbalance faults of a blower has been studied in Knowledge Discovery Laboratory at Norwegian University of Science and Technology. Experimental results on the blower show the effectiveness and robustness in diagnosing and predicting the faults using proposed methods. This paper is organized as the following: Section 2 introduces the data acquisition experiment setup and procedure. Section 3 describes the procedure of proposed methods and the following Section 4, Section 5, and Section 6 introduce the principle of feature extraction using wavelet packets transform, principal component analysis, and BP neural networks, respectively. A case study for validating the method is presented in Section 7, and finally the conclusions and discussions are given in Section 8.

2 Data acquisition experiment

To research how to apply intelligent fault diagnosis and prognosis framework to monitor equipment or machines, a simple experimental setup is established in Knowledge Discovery Laboratory.

2.1 Experimental setup

Figure 1 shows the hardware of the experimental setup which includes a blower, three vibration sensors, power supply for sensors, connector, DAQ card, and a computer. In this setup, the blower is selected as our monitoring object and a kind of vibration sensors (Kistler, Type 8702B100) are chosen to collect the signals from the blower. Three sensors are setup on the blower in three directions which can collect the vibration signals in different directions (Fig. 2). The signals are collected from the sensors and processed using some processing method like filter, de-noising, and compression. Then, the features are extracted in wavelet domain which can be used to train and query BP network. After training, the system can judge the real states of monitored components using real time signals.

2.2 Experimental procedure

In the present study, four different degradation levels of unbalance are simulated using three different parts (Fig. 3), which are mounted in the axis end of the blower. The unbalance degradation (condition) contains 0, 0.3, 0.7, and 1 which represents the performance states from perfect to absolutely failure (unbalance). In the first case, power on the blower and collect and store signals from the sensors without mounting any simulation part. Next, power off the blower and mount first part in the axis end, and then, power on the blower and collect and store the signals from sensors. Repeat this process until all the degrading signals simulated by simulation parts were collected. Figure 4 shows the signals of the second sensor from perfect state to absolutely failure.

3 The procedure of fault diagnosis and prognosis

The pattern classification theory has been a key factor in fault diagnosis and prognosis methods development. Some classification methods for equipment performance monitoring use the relationship between the type of fault and a set of patterns which is extract from the collected signals without establishing explicit models. Currently, ANN is one of the most popular methods in this domain. ANN is a model that emulates a biological neural network [27]. The origin of ANN can be traced back to a seminar paper by McCulloch and Pitts [28] that demonstrated a collection of connected processors,

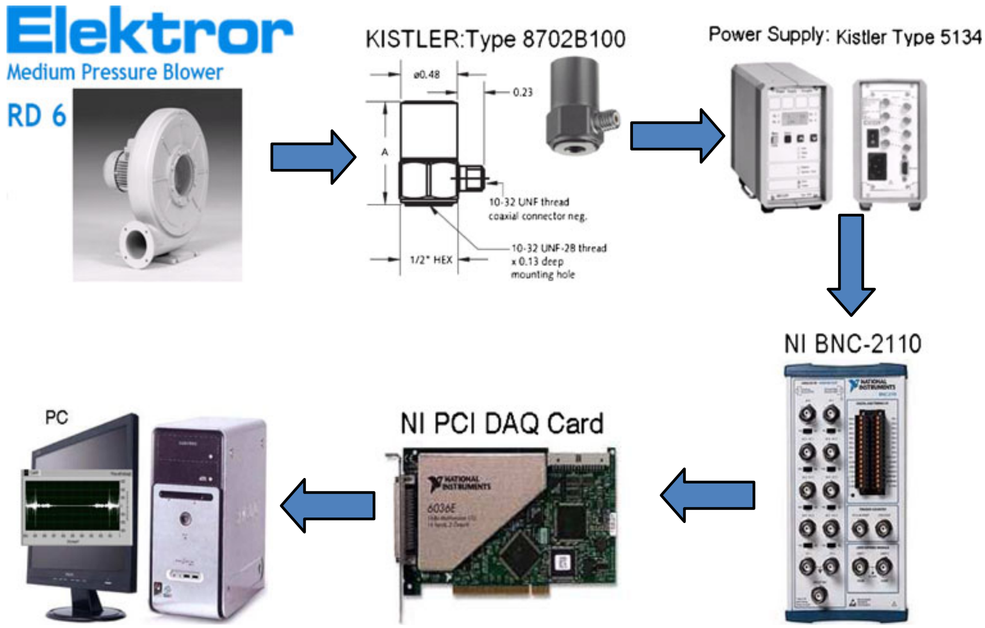


Fig. 1 Hardware of experimental setup

loosely modeled on the organization of brain, could theoretically perform any logical or arithmetic operation. Then, the development of ANN techniques is very fast which is extensive to many categories containing back propagation (BP), self-organization mapping, radial basis function, etc. The application of artificial neural network models lies in the fact that they can be used to infer a function from observations. This is particularly useful in applications where the complexity of the data or task makes the design of such a function by hand impractical. This attribution is very nontrivial in diagnostic problems. BP network is a main type of ANN used to solve fault diagnosis and prognosis problem. This section mainly describes how to integrate BP network, PCA, and wavelet transform (WT) for pattern classification in fault diagnosis and prognosis, and the following three sections will introduce the principle of these three techniques.

The PCA new features generated from the features in wavelet domain (standard deviation of wavelet packet coefficients (SDWPC)) of vibration signals are used to estimate the fault status of components and machines. The input nodes of BP neural network come from the test signals sensors. BP neural network was made up of one input layer, one output layer, and one hidden layer of nodes. And it has been proved that such three-layer BP neural network model can approach any continuous functions at any precision. The values of output are from 0 to 1 which represent from perfect condition to complete failure of specific fault.

Because of convenience of handling the signal collection, signal processing, and interface things, the Labview is selected as program software in this project. However, the capability of mathematical calculation of Labview is not as good as Matlab. Therefore, both software are combined to apply in this project. The procedure of fault diagnosis and prognosis integrating BP network, PCA, and WPC is shown in Fig. 5. The historic data are collected and processed which are fist two steps. Then, the features in wavelet domain (SDWPC) are extracted from the processed signals. These features are analyzed by the PCA which can generate new features called principal component which can used to train ANN. After training, the signals in real time are collected and used to query the BP network, and then the condition of the monitored components can be obtained.

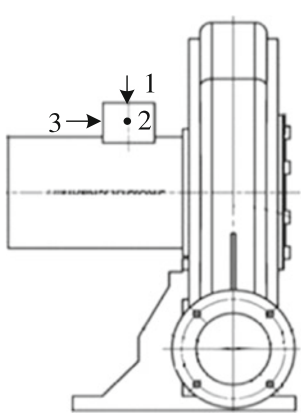


Fig. 2 Sensors setup on blower

Fig. 3 Parts for simulation degradations

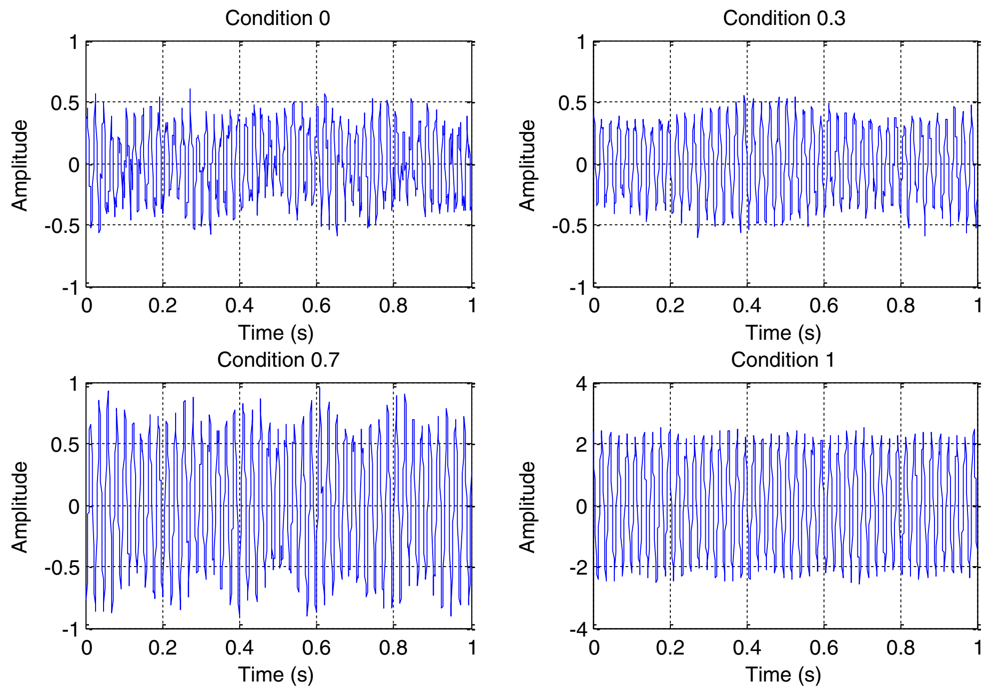


Fig. 4 Raw signals with different degradations

4 Features extraction in wavelet domain

To monitor the conditions of manufacturing equipments, collecting the data from them through suitable sensors is the first step which was described above. Then, the collected signals should be processed using many methods of signal processing such as filtering, amplification, data compression, data validation, and de-nosing which generally aim is improving the signal-to-noise ratio. After that, the features should be extracted from processed signals which can be used to decide whether the performance is perfect or degrading and decide the remaining useful life of monitored components. Finally, the maintenance decision can be made according to the status of the equipment. We can see from the above process that feature selection has a significant impact on the success of fault diagnosis and prognosis and making the correct maintenance decision. The features can be extracted from the processed signals in time domain, frequency domain, time–frequency domain, and wavelet domain which should be fairly insensitive to noise and within fault class variation. This part will introduce extraction features from wavelet domain briefly.

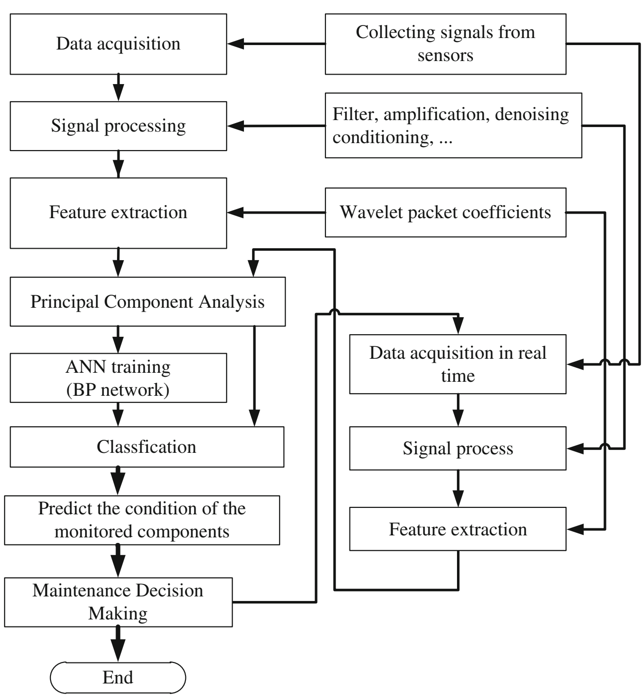


Fig. 5 The procedure of fault diagnosis and prognosis integrating BP network, PCA, and WT

WT is a time–frequency decomposition of a signal into a set of "wavelet" basic function. Wavelet analysis has proved its great capabilities in decomposing, de-noising, and signal analysis which made the analysis of non-stationary signals achievable as well as detecting transient feature components as other methods were inept to perform since wavelet can concurrently impart time and frequency structures. WT gives good time and poor frequency resolution at high frequencies, and good frequency and poor time resolution at low frequencies. Analysis with wavelets involves with breaking up a signal into shifted and scaled versions of the original (or mother) wavelet, i.e., one high-frequency term from each level and one low frequency residual from the last level of decomposition.

Continuous wavelet transform [29] is a function of two parameters, explicitly time and frequency and, therefore, bears a great information redundancy in signal or function being analyzed. Discretized wavelet transform (DWT) [30]

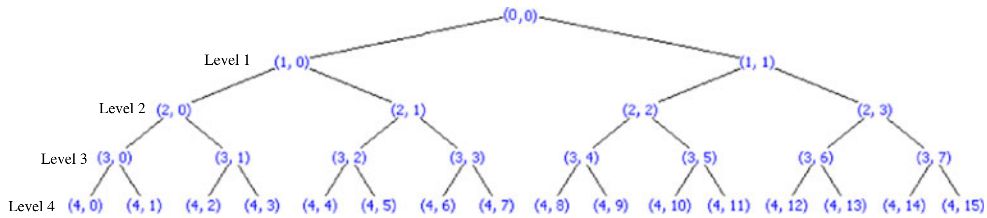


Fig. 6 Structure of wavelet packet transform (four levels)

analysis is more efficient still with the identical accuracy. In order to shift along the time axis, DWT analyzes the signal by applying a wavelet filter with a specific frequency band which is dependent on the level of decomposition. Consequently, the signal can be decomposed into a hierarchical structure with wavelet details and approximations at various levels as follows:

$$f(t) = \sum_{i=1}^{i=j} D_i(t) + A_j(t) \tag{1}$$

Where $D_i(t)$ denotes the wavelet detail and $A_j(t)$ stands for the wavelet approximation at the jth level.

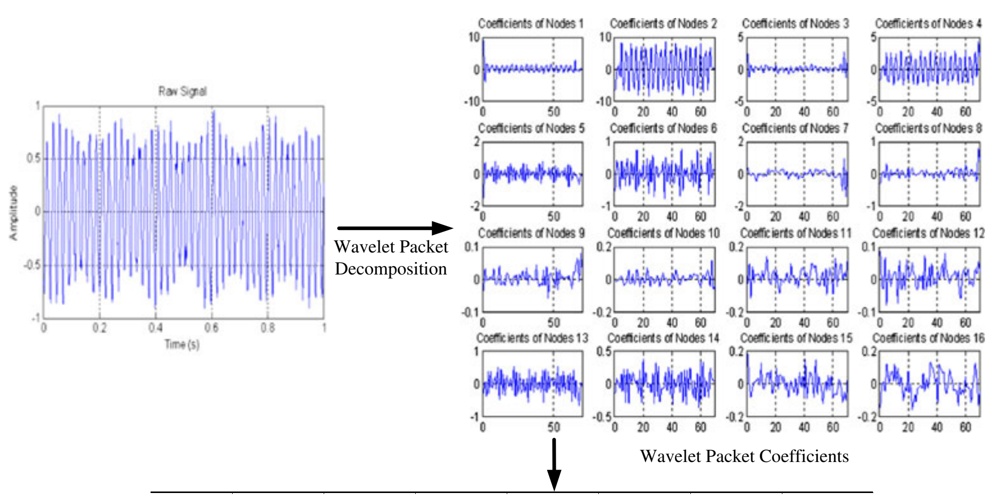
Wavelet packet method [31] which is a generalization of wavelet decomposition offers a richer range of possibilities for signal analysis. Contrary to WT, the wavelet packets contain a complete set of decompositions and details at every level and hence providing a higher resolution in the high-frequency region, i.e., the wavelet detail component at each level is further decomposed to obtain its approximation and detail components. The structure of wavelet packet decomposition (WPD) algorithm that broke up to four resolution levels is shown in Fig. 6. In the figure, the node (4, 0) presents the symbol for a subspace that stands for the fourth resolution and the 0th subspace. For this case, each node presents the frequency bandwidth 64 Hz which means that a node (4, 0) presents the signal character of the bandwidth between 0 and 64 Hz.

Since different types of wavelet functions have different time–frequency structures, a function with a time–frequency structure matching superlatively that of the transient component must be used to effectively detect the transient component. In general, the smooth wavelets are better for regular, stationary, periodic data and the compact wavelets are better for non-stationary, transient data [32]. As a result, Daubechies 4 wavelet function has been chosen for this case after several trials as it is often chosen arbitrarily for signal analysis and synthesis by experiments in many papers in the field (e.g., [33]) and there is no computational logic behind the selection of Daubechies order.

It is observed that the scaling function has a low-pass form, whereas the wavelet function has a high-pass form. Thus, the wavelet function is essentially responsible for extracting the detail (high-frequency components) of the original signal.

For each processed signal, wavelet packet was applied up to the fourth level, thus giving 16 signal coefficients. The wavelet packet coefficients and their corresponding standard deviations for one signal are shown in Fig. 7. In the end, the SDWPC of the processed signals is selected as feature

Fig. 7 Wavelet packet coefficients (WPC) and their relevant standard deviation



Node 1	Node 2	Node 3	Node 4	Node 5	Node 6	Node 7	Node 8
1.575	5.452	0.631	1.924	0.443	0.395	0.302	0.15
Node 9	Node 10	Node 11	Node 12	Node 13	Node 14	Node 15	Node 16
0.024	0.032	0.055	0.034	0.316	0.194	0.063	0.071

Standard Deviation

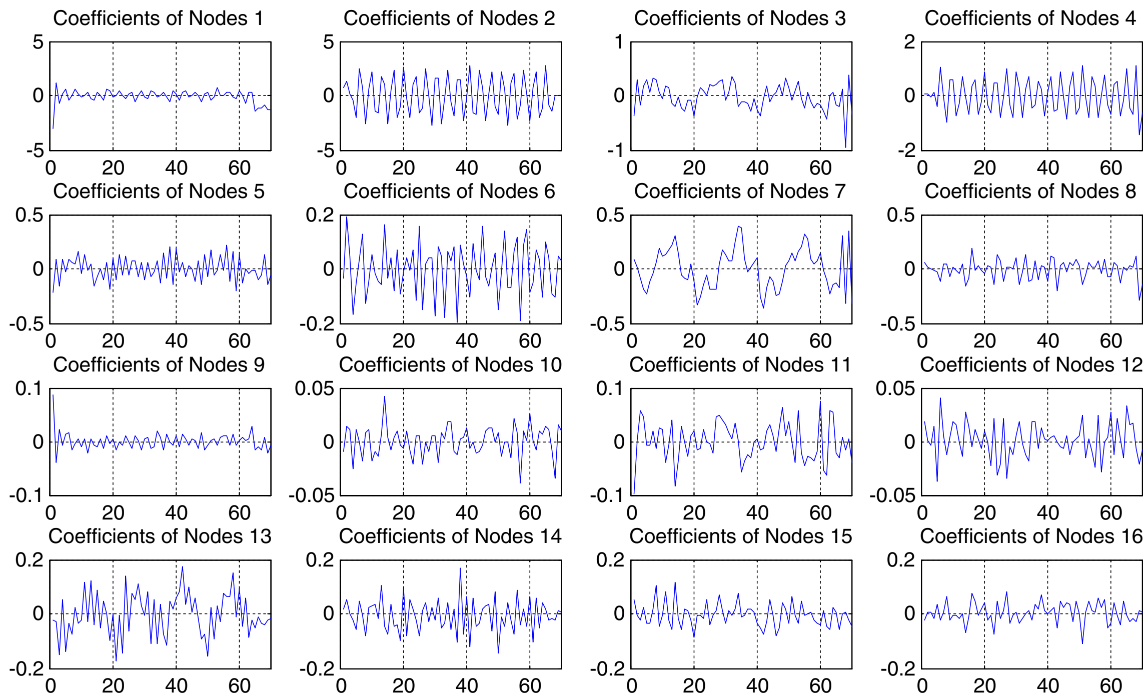


Fig. 8 Wavelet packet decomposition of condition 0.7

vector which is used to train ANN after PCA analysis. The decomposition using wavelet packet method for signals of degradation 0.7 is shown in Fig. 8.

5 Principal component analysis

The concept of PCA was invented in 1901 by Karl Pearson [34]. It is a mathematical procedure that uses an orthogonal transform to convert a set of observations of possibly correlated variables into a set of values of uncorrelated variables called principal components. This transform is defined in such a way that the first principal component has as high a variance as possible, which means accounting for as much of the variability in the data as possible, and each succeeding component in turn has the highest variance possible under the constraint that it be uncorrelated with the preceding components. It can reduce data dimension and eliminate multi-collinearity. Currently, PCA is mostly used to reduce the dimension while maintaining the main information in data mining analysis and making models. This part will introduce PCA briefly.

PCA computes a new set of uncorrelated multivariate (vector) samples by a transform of coordinate rotation from original correlated multivariate samples. In our case, a matrix composed by n rows which means n samples are collected and m (m=16) columns which represent the number of features are expressed as below:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \tag{2}$$

PCA can obtain a new set of vector according to the following steps:

1. Calculate the correlation coefficient matrix

   The correlation coefficient matrix is calculated according to the following equation:

$$\begin{aligned} R &= \mathrm{Cor(i,j)} \\ &= \frac{(\mathrm{n}-1)\cdot \mathrm{Cov(i,j)}}{\sqrt{\sum_{\mathrm{k}=1}^{\mathrm{n}} (\mathrm{x_i(k)}-\mu_\mathrm{i})^2 \sum_{\mathrm{k}=1}^{\mathrm{n}} \left(\mathrm{x_j(k)}-\mu_\mathrm{j}\right)^2}} \\ &= \left(\mathrm{r_{ij}}\right)_{\mathrm{m\times n}} \end{aligned} \tag{3}$$

   where n is the number of samples. The dimension of the correlation matrix R is $m\times n$. Cov(i, j) means the

Table 1 Variance for each component

Component no.	1	2	3	4	5	7	7	8	9	10	…
Variance	5625.841	0.681	0.424	0.133	0.053	0.005	0.002	0.002	0.001	0	…

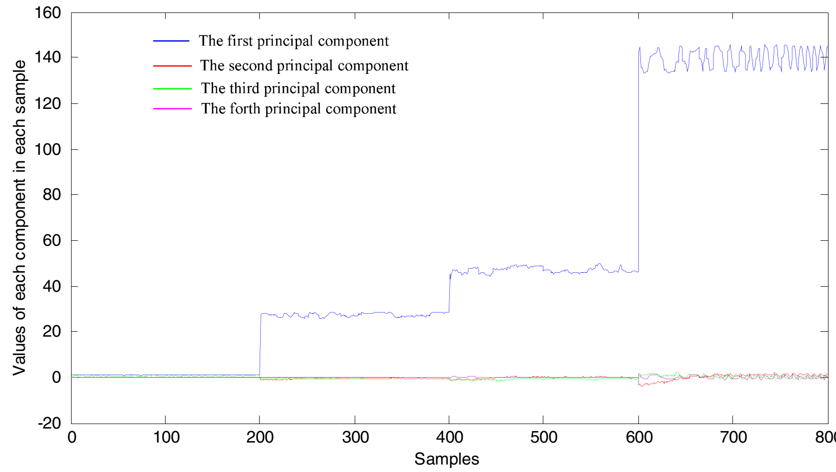


Fig. 9 First four principal components

covariance which matrix is $m \times n$ and can be expressed as:

$$\mathrm{Cov}(\mathrm{i},\mathrm{j}) = \frac{1}{(\mathrm{n}-1)}(\mathrm{x_i} - \mu_\mathrm{i})\left(\mathrm{x_j} - \mu_\mathrm{j}\right) \quad i,j = 1,2,\ldots,m \tag{4}$$

where μ_i and μ_j are the averages of the ith and jth rows of matrix $\boldsymbol{X}$, respectively.

2. Calculate the eigenvectors and eigenvalues of the matrix R

   The m eigenvalues λ_i which have the constraint as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ and their responding eigenvectors $\boldsymbol{V_i}$ are calculated from correlation matrix. λ_i and $\boldsymbol{V_i}$ satisfy the following equation:

$$AV_i = \lambda_i \cdot V_i \quad i = 1,2,\cdots,m \tag{5}$$

   where A is a $m \times n$ covariance matrix or correlation matrix and the vector $\boldsymbol{V_i}$ can be expressed as $V_i = [V_{1i}, V_{2i}, \cdots, V_{mi}]$.

3. Generates the new samples

   A new set of uncorrelated multivariate (vector) samples are computed according to the following equation:

$$X_{\mathrm{new}} = V^T \cdot X \tag{6}$$

   where X_{new} is the new uncorrelated multivariate (vector) samples, and $\boldsymbol{X}$ is the original correlated multivariate (vector) samples. Both of them are $n \times m$ matrices whose row vectors represent a single channel sample. $\boldsymbol{V}$ is eigenvectors matrix which is also called the weight matrix. Each column of $\boldsymbol{V}$ is one principle component. X_{new} is the principal component scores. Each row of X_{new} is the scores for one principal component. Each λ_i is variance of the scores for one principal component. Most of time, only first several components in X_{new} are selected as principal components according to the variance threshold.

In our experiment, 200 samples for each condition were collected as training data and were analyzed by PCA. There are 48 variables in each sample from part 3. Now, the original sample matrix's dimension is 800×48. Then, these data are analyzed by PCA. The variance for each component is shown in Table 1 (only the first ten values are shown) and the first four principal components were displayed in Fig. 9. If the value of threshold is set to ε=1, only the first principal component was selected as feature to train ANN. If the value of threshold is set to ε=0.5, only the first two principal components were selected as features to train ANN.

6 BP neural network

BP neural network which is the most widely used neural network model currently was proposed by Rumelhart and McCelland in 1986 [35]. It is a multilayer feed-forward network usually containing the input layer, hidden layer, and output layer (Fig. 10), which trained by an error back propagation algorithm. The biggest advantage of ANNs trained by back propagation is that there is no need to know

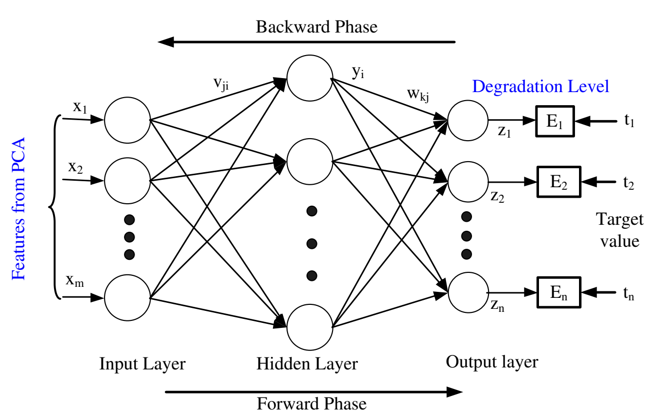


Fig. 10 BP neural network with single hidden layer

the exact form of analytical function on which model should be built. So it is not necessary have neither the function type not even the number and position of the parameters in the model function. Moreover, BP network can learn and store a lot of input–output model mapping without mathematical equations which describing this mapping. The learning method of BP is the steepest descent method which is adjusting the weights and thresholds of the network to minimize the sum of squared errors. The general procedure of BP network training can be summarized as follows [27]:

1. Initialize the weights to small random vales (−1, 1);
2. Select a training vector pair (input and the corresponding desired output) from the training set and present the input vector to the inputs layer of the ANN;
3. Calculate the actual outputs (forward phase);
4. Adjust the weights ω_{ji} to reduce the difference according to the error between actual output and target (backward phase);
5. Return to step 2 and repeat for each pattern p until the total error has reached an acceptable level;
6. Stop.

Figure 10 shows a BP network structure with a single hidden layer. $\overrightarrow{x}$ and $\overrightarrow{t}$ are input and target of training data, respectively. ν_{ji} and ω_{kj} are weights between input and hidden layer, and between output and hidden layer, respectively. y_i and z_k are outputs of hidden and actual output of output layer. The objective of ANN training is to obtain all the suitable weights to meet the input and the target of training data. After the training of BP network, for each set of test data or query data, there is a set of output calculated by the final updated weights. For a specific application in fault diagnosis and prognosis, after training by features extracted from processed historic data, the BP network can classify the fault and predict the states of the monitored components or machine units. In this case, the input features are new vectors (features) from PCA of SDWPC while the output means the degradation level for each equipment or component.

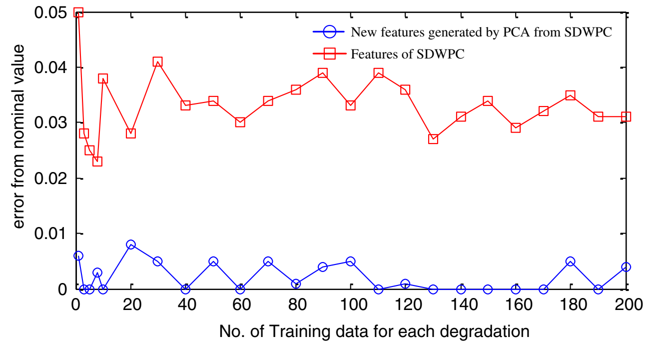


Fig. 12 Errors of condition 0

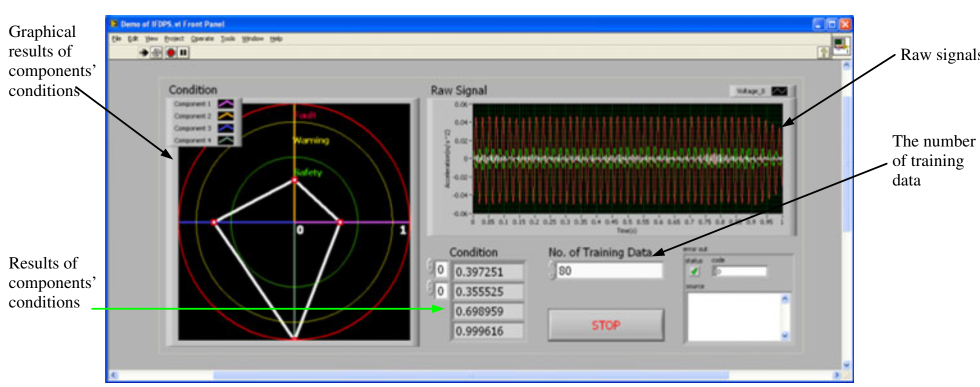
7 Case study

A framework called Intelligent Blower Fault Diagnosis and Prognosis System is developed in Knowledge Discovery Laboratory to show how to apply the proposed methods in real system. This framework is a part of SFI-Norman project called condition-based maintenance in order to achieve near-zero breakdown manufacturing and further to reach zero defect manufacturing. In this part, some experiments are done to certify the correctness, robustness, and precision of proposed methods and comparison of the different result among using different inputs.

The hardware of the framework is in the same as in Fig. 1 and monitored component is the same as in Fig. 2. Figure 11 shows the real time condition of the blower. The raw signals that are collected from sensors in real time are displayed on the top right of the interface. The number of training data can be adjusted by the user. After training, the conditions of the monitored components are calculated by ANN. The results can be displayed graphically as the figure on the left. The area inside the blue circle represent safe condition, the area between blue circle and yellow circle represent warning condition, while the area between yellow circle and red circle represent failure condition. The points located in radial line represent the conditions of the specific components.

In this case study, four conditions are defined for the monitored component which are 0, 0.3, 0.7, and 1. They

Fig. 11 Interface of system



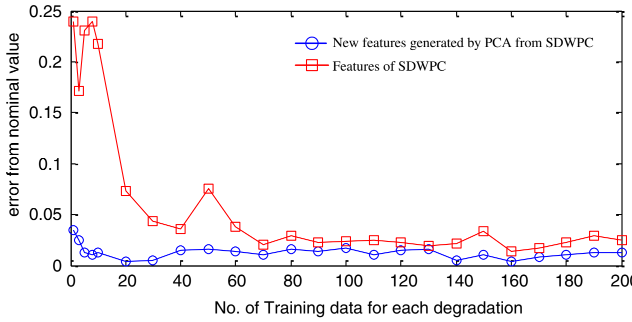


Fig. 13 Errors of condition 0.3

represent from perfect performance (condition 0) to completely failure (condition 1) discretely. For each condition, 200 training signals are collected and processed. The new feature vectors are generated using PCA from SDWPC. These new features are put into ANN for training. Finally, test signals are collected and processed like the training data. In this experiment, for each condition, 20 samples are collected which were used to test trained ANN for verification.

For each testing data, the output of ANN and the nominal values which can be called “error from nominal value” (average value of testing data for each condition) are compared. The values of these errors are shown in Figs. 12, 13, 14, and 15. There are two curve lines for each figure. One represents only using the features of SDWPC as inputs of ANN. The other represents using the new features generated by PCA from the features of SDWPC as inputs of ANN.

All these four figures show the differences between the predicted values and nominal values of the four different conditions using the features of SDWPC and new features generated by PCA from SDWPC. Figure 12 shows the result of condition 0. The error is much smaller of the result using the new features generated by PCA from SDWPC compared to using the features of SDWPC as inputs of ANN. Figure 13 and Fig. 15 show the results of condition 0.3 and condition 1, respectively. When the number of training sets is very small, the results using new features generated by PCA from SDWPC are much better than using features of SDWPC in these two figures. However, with the number of the training data increasing, the results of using both

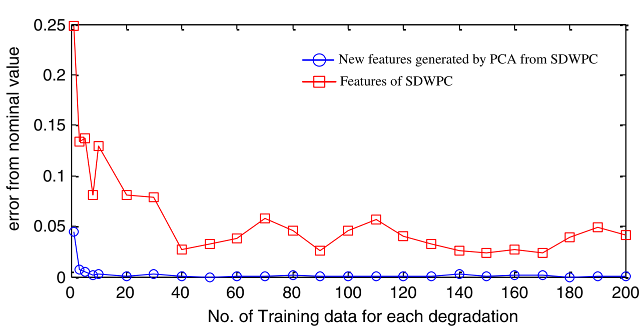


Fig. 14 Errors of condition 0.7

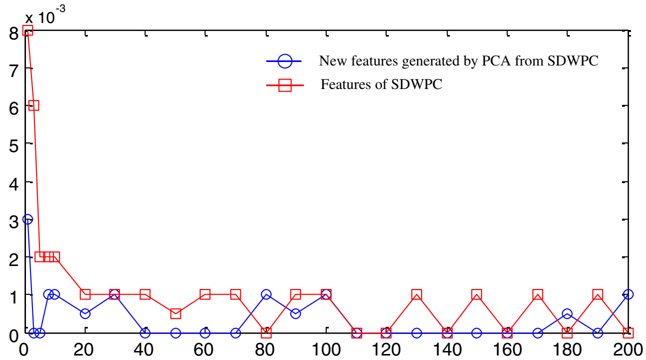


Fig. 15 Errors of condition 1

features are almost the same in these two figures and both of them are correct and precise. Figure 14 shows the result of condition 0.7. In this figure, in both kinds of features, the performance is very effective and corrective whatever the number of training data is, but the result of using the new features generated by PCA from SDWPC is much better than using features of SDWPC. As what we can see from Fig. 13 to Fig. 14, when the condition is neither perfect nor completely failure, the result of using SDWPC is not believable if the number of training data is very small because the “error from nominal value” is large. But it is still believable of using new features generated by PCA from SDWPC to training and testing ANN in these conditions. We can see from the four pictures that the precision is better using new features generated by PCA from SDWPC than using features of SDWPC in any condition and in any number of training data.

8 Conclusions and discussions

In this paper, a new method integrating wavelet transform, principal component analysis, and BP ANN for fault diagnosis and prognosis was proposed. To verify the correctness and effectiveness of this method, a framework called Intelligent Blower Fault Diagnosis and Prognosis System was established as a case study. This method demonstrated high effectiveness in prognosing machine faults. It can predict the degradation and condition of the monitored components. PCA was applied in this method to reduce the input dimension (number of variables) of ANN without omitting the useful information. ANN model may become over specified, i.e., more input variables than is strictly necessary, due to including superfluous variables which are uninformative, weakly informative, or redundant [36]. In this case, the total volume of the modeling problem domain increases exponentially with the linearly increasing of variable dimensionality which is called curse of dimensionality [37]. This will cause many problems such as:

computational burden increasing which is a significant influence in determining speed of training and training difficulty due to inclusion of redundant and irrelevant input variables. By reducing the dimensionality of variables, PCA can solve these problems and improve the effectiveness of ANN training. Therefore, the proposed method provides a faster, more effective, and more precise solution for fault diagnosis and prognosis.

In this paper, the minimum bandwidth 0~64 Hz is chosen in WPD because the fundamental frequency of the vibration signal is 47.5 Hz. In a real system, the minimum bandwidth of WPD (which means how many levels should be decomposed) should be selected according to the real fundamental frequency. In the case study of this paper, there is only one type fault (unbalance) simulated. In the future, multi-fault diagnosis and prognosis should be a research topic. The proposed methods can be applied to decide many other faults such as wear, crack, and fatigue of bearings and gearbox which faults can be reflected by vibration signals. To apply this method, the fundamental frequency has to be known firstly and thereafter the sample rate of vibration signals, the level of wavelet decomposition, and the structure of BP network can be determined properly. The degradation information could be very useful for maintenance decision making, and thus, how to apply this degradation information in maintenance decision making should be a research issue as well in the future.

Acknowledgments This paper is a result of Norwegian Manufacturing Future project in Center of Research and Innovation of Norwegian Manufacturing Future (SFI-Norman), which is financially supported by Norwegian Research Council.

References

1. Hu W, Starr AG, Zhou Z, Leung AYT (2000) A systematic approach to integrated fault diagnosis of flexible manufacturing systems. Int J Mach Tool Manuf 40:1587–1560
2. Kegg RL (1984) On-line machine and process diagnostics. Annals of the CIRP 32(2):469–573
3. Zhou ZD, Chen YP, Fuh JH, Nee AYC (2000) Integrated condition monitoring and fault diagnosis for modern manufacturing system. Annals of CIRP 49(1):387–390
4. Lee J, Ni J, Dragan D, Qiu H, Liao H (2006) Intelligent prognostics tools and e-maintenance. Comput Ind 57:476–489
5. Ripka P, Tipek A (2007) Modern sensors handbook. ISTE, London
6. Samanta B, Al-Balushi KR, Araimi SA (2003) Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. Eng Appl Artif Intell 16:657–665
7. Ge M, Du R, Zhang G, Xu Y (2004) Fault diagnosis using support vector machine with an application in sheet metal stamping operations. Mech Syst Signal Process 18:143–159
8. Zhang L, Nandi AK (2007) Fault classification using genetic programming. Mech Syst Signal Process 21:1273–1284
9. Lei Y, He Z, Zi Y, Hu Q (2007) Fault diagnosis of rotating machinery based on multiple ANFIS combination with gas. Mech Syst Signal Process 21:2280–2294
10. Goebel M, Gruenwald L (1999) A survey of data mining and knowledge discovery software tools. ACM SIGKDD 1:20–33
11. Kasabov N (2001) Evolving fuzzy neural networks for supervised/unsupervised online knowledge-based learning. IEEE Trans Syst Man Cybern B Cybern 31:902–918
12. Marzi H (2004) Real-time fault detection and isolation in industrial machines using learning vector quantization. Proc Inst Mech Eng, Part B: J Eng Manuf 218:949–959
13. Markou M, Singh S (2003) Novelty detection: a review—part 2: neural network based approaches. Signal Process 83:2499–2521
14. Wang K (2002) Intelligent condition monitoring and diagnosis systems. IOS, Amsterdam
15. Sreejith, B., Verma A.K. and Srividya, A. (2008). Fault diagnosis of rolling element bearing using time-domain features and neural networks. The Third International Conference on Industrial and Information Systems, pp. 1–6
16. Iyananda H, Towill DR (1973) Fault diagnosis using time domain measurements. Radio and Electron Engineer 43(9):523–533
17. Chang J, Li T, Li P (2010) The selection of time domain characteristic parameters of rotating machinery fault diagnosis. 2010 Int Conf Logist Syst Intell Manag 1:619–623
18. Brygilewicz V, Wojciechowski J (1998) Time domain fault diagnosis of analogue circuits in the presence of noise. IEE Proc Circ, Devices and Syst 145(2):125–131
19. Chang H (2002) Frequency–domain grouping robust fault diagnosis for analog circuits with uncertainties. Int J Circ Theory and Appl 31(1):65–86
20. Tian X, Lin J, Ken RF, Zuo MJ (2003) Gearbox fault diagnosis using independent component analysis in the frequency domain and wavelet filtering. 2003 IEEE Inter Conf Acoust, Signal Proc 2:245–248
21. Purkait P, Chakravorti S (2002) Time and frequency domain analyses based expert system for impulse fault diagnosis in transformers. IEEE Trans Dielectr Electr Insul 9(3):433–445
22. Wu J, Liu C (2009) An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network. Expert Syst Appl 36:4278–4286
23. Kamal H, Meisam P, Hosein M (2011) Fault diagnosis and classification based on wavelet transform and neural network. Progress in Nucl Energy 53:41–47
24. Saravanan N, Ramachandran KI (2010) Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN). Expert Syst Appl 37:4168–4181
25. Qiu, C., Zuo, X., Wang C., Wu J., and Zhang, T. (2010). An urban traffic speed fusion method based on principal component analysis and neural network. The 2010 International Joint Conference on Neural Networks, pp. 1–7.
26. Zhu, Z., Ma, Z., Wang, Z. and Jiang, J., (2009). Model study of transformer fault diagnosis based on principal component analysis and neural network. Proceeding of 2009 IEEE International Conference on Networking, Sensing and Control, pp. 936–940
27. Wang K (2005) Applied computational intelligence in intelligent manufacturing systems. Advanced Knowledge International Pty Ltd, Australia
28. McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5(4):115–133. doi:10.1007/BF02478259
29. Soman KP, Ramachandran KI (2004) Insight into wavelets from theory to practice, 2nd edn. PHI Learning Private Ltd, Delhi
30. Goumas S, Zervakis M, Pouliezos A, Stavrakakis GS (2001) Intelligent on-line quality control of washing machines using

discrete wavelet analysis features and likelihood classification. Eng Appl Artif Intell 14(5):655–666
31. Li X, Qu L, Wen G, Li C (2003) Application of wavelet packet analysis for detection in electro-mechanical systems based on torsional vibration measurement. Mech Syst Signal Proc 17 (6):1219–1235. doi:10.1006/mssp.2002.1517
32. Staszewski WJ, Giacomin J (1997) Application of the wavelet based FRFs to the analysis of non-stationary vehicle data. Proc Int Modal Anal Conf 1:425–431
33. Vafaei S, Rahnejat H (2003) Indicated repeatable runout with wavelet decomposition (IRR-WD) for effective determination of bearing-induced vibration. J Sound and Vib 260(1):67–82. doi:10.1016/S0022-460X(02)00900-8
34. Pearson K (1901) On lines and planes of closest fit to systems of points in space. Philos Mag 2(6):559–572
35. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representations by error propagation. In: Rumenhart DE, McCelland JL (eds) Parallel distributed processing: explorations in the microstructure of cognition. MIT, Cambridge, pp 318–362
36. May, R., Dandy, G., and Maier, H. (2011). Review of input variable selection methods for artificial neural networks. In: K. Suzuki (ed.) Artificial neural networks—methodological advances and biomedical applications. InTech, Croatia. doi:10.5772/16004
37. Bellman R (1961) Adaptive control processes: a guided tour. Princeton University Press, New Jersey