IndoMS Journal on Industrial and Applied Mathematics
Volume. 2, Issue. 1 (2015), pp. 41-48

THE DOMINANT FEATURES INFLUENCE IN CLASSIFICATION OF BEARING CONDITION USING PRINCIPAL COMPONENT ANALYSIS AND GREY RELATIONAL ANALYSIS

LILI A. WULANDHARI *

Abstract. Condition classification in bearing case study involves a large number of features to obtain an accurate classification. However, a large number of features need high processing time for classification. Therefore, it is a requirement to determine the dominant features which has high influence in classifying the condition of the bearings. This paper propose a techniques to find the dominant features of the bearing using principal component analysis and grey relational analysis, while the classification of the bearing condition use back propagation neural network. The experimental result shows that generally PCA-BPNN has better performance with four dominant features compared with GRA-BPNN which use five dominant features.

Key words. Back Propagation Neural Network, Classification, dominant features, grey relational analysis, principal component analysis

1. Introduction. Condition classification in a bearing is one of the important requirements in maintenance task of industry. A bearing is an equipment that prevent relative motion between two moving parts and reduce friction on rotating shaft by providing metal ball or roller. They are widely used in many applications and different applications have different kind of bearing used. For example the tapered roller bearings are used for automobile wheels (as shown in Figure 1.1 [5]), the cylindrical roller bearing for aircraft GA turbine engine, and needle roller bearing for car follower assembly [4]. Appropriate bearing designs can minimize the friction and its failure may cause expensive loss of production [3]. However, bearing is one of machine parts which has a high percentage of defect compared to the other components [16].

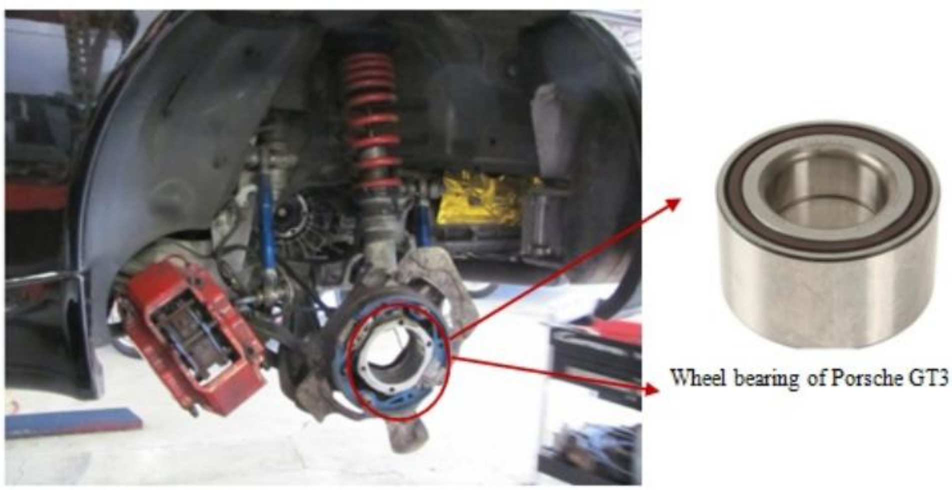

FIG. 1.1. *Example of tapered roller bearing in automobile wheels*

Therefore, it is important to find a technique to classify the condition of the bearing accurately. Back propagation is one of a good technique to classify the condition of the bearing [6], it is used to model the behaviours of the system which are then classified. BPNN is an suitable tool for modelling the behaviours of a system since they have the following three important characteristics: generalization ability, noise tolerance and fast response once trained [14]. Even if the training data are affected by noise, BPNN will still be able to generalize the system behaviour with the level of accuracy being proportional to the level of noise [1]. However, for large number of features such as in bearing case involves the BPNN complexity, consequentlyaffect to the processing time to obatain good accuracy [8]. In order to reduce the BPNN complexity, the dominant features for condition classification must be determined and chosen correctly since choosing the features randomly to be used as inputs will consequently influence

*School of Computer Science, Bina Nusantara University, Jl. K.H. Syahdan No.9 Jakarta, Indonesia, 11480 (lwulandhari@binus.edu).

the classification accuracy and become time consuming [2]. The dominant features are the features that contain the most useful information regarding to the bearing condition. By finding these features an accurate classification of bearing condition can be obtained in less complexity and processing time.

This paper proposes technique of classification by applying PCA or GRA to determine the dominant features from the bearings. The experiments are executed to the PCA-BPNN and GRA-BPNN, the result will be compared and analyzed to see which techniques between PCA-BPNN and GRA-BPNN has the better result in condition classification. This paper is arranged into four sections; first section is introduction, and then followed by the basic theory about PCA, GRA and BPNN. The third and fourth sections contain the proposed methodology and experimental result and analysis respectively. Finally the paper is ended by the conclusion in the fourth section.

2. Research Methodology. This paper contains three basic concepts to achieve good accuracy namely grey relational analysis (GRA), principal component analysis (PCA) and back propagation neural networks (BPNN). GRA and PCA are used to determine the dominant features to be used in BPNN for condition classification. The explanation of these concepts is presented in the next section.

2.1. Grey Relational Analysis. GRA utilizes the mathematical method to analyzing correlation between the references series which is the ideal value of features and the alternatives series [9]. It firstly normalizes the features and then translates the performance of all alternatives into a comparability sequence with the ideal value called grey relational generating [7], followed by the calculation of grey relational coefficient between all comparability sequences and the references sequences. Finally, the grey relational grade between the reference sequence and every comparability sequence is calculated based on the grey relational coefficient. The highest grey relational grade of the alternatives features indicates that the features have dominant influence to the condition classification. The procedures of GRA are shown in Figure 2.1.

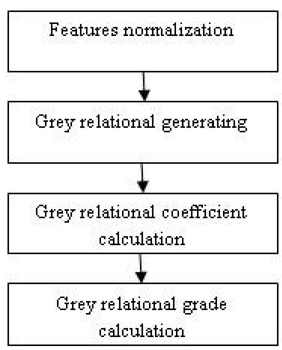

FIG. 2.1. *Grey Relational Analysis Procedures*

The GRA starts by normalizing the features, let x_{ij} is the ith sample of the jth features where $i = 1, 2, ..., S$ and $j = 1, 2, ..., F$

$$\begin{aligned} x_{max} &= max(x_{ij}|i = 1, 2, ..., S; j = 1, 2, ..., F) \\ x_{min} &= min(x_{ij}|i = 1, 2, ..., S; j = 1, 2, ..., F) \end{aligned} \tag{2.1}$$

Then the normalized features x^*_{ij} can be obtained by equation as follows [11]:

$$x^*_{ij} = \frac{x_{ij} - x_{min}}{x_{max} - x_{min}} \tag{2.2}$$

Next, Grey relational generating is conducted by determining the reference or ideal values of the features. Let P_j is the features, with $j = 1, 2, ..., F$. C_s is the condition of the bearing systems with $s = 1, 2, ..., C$. Ideal values of the features z_{ij} are the average of sth features value in jth condition. If in a condition class there are fifteen samples, then the ideal values of the features can be written as follows:

$$z_{xj} = \frac{1}{15} \sum_{i=15s-14}^{15s} x^*_{ij} \tag{2.3}$$

Regarding to obtain the next step namely grey relational coefficient, I need to determine the comparability sequences of the alternatives values and the ideal values. In this paper, I have features $F = 10$ and $S = 240$ samples which consist of 15 samples for 16 condition classes. So that in 240 samples I can define that

$$\begin{array}{llll} w_{ij} = z_{1j} & \text{for} & j = 1, ..., 10; & i = 1, ..., 15 \\ w_{ij} = z_{2j} & \text{for} & j = 1, ..., 10; & i = 16, ..., 30 \\ w_{ij} = z_{3j} & \text{for} & j = 1, ..., 10; & i = 31, ..., 45 \\ w_{ij} = z_{nj} & \text{for} & j = 1, ..., 10; & i = 15(n-1), ..., 15n \end{array} \tag{2.4}$$

where $n = 1, ..., 16$ then the comparability of the alternatives and ideal values can be calculated as

$$\Delta w_{ij} = |x^*_{ij} - w_{ij}| \tag{2.5}$$

where Δw_{ij} is the comparability of alternatives and ideal values, x^*_{ij} is the alternatives value which is normalized features of the vibration signal data and w_{ij} is the ideal values which is defined based on the condition classes. Based 2.4 and 2.5, the next step of GRA procedures namely grey relational coefficient calculation between comparability sequences and ideal sequences is written as follows

$$GRC_{ij} = \frac{min_k(min_l(\Delta w_{kl})) + \xi max_k(maxk(\Delta w_{kl}))}{(\Delta w_{ij}) + \xi max_k(maxk(\Delta w_{kl}))} \tag{2.6}$$

for $k = 1, 2, ..., 240; i = 1, 2, ..., 10$

where, GRC_{ij} is the grey relational coefficient value of ith samples and jth feature, ξ is the distinguishing coefficient which is defined in the range $0 \leq \xi \leq 1$. Then, the grey relational grade (GRG) is determined by averaging the GRC to each feature and represented as equation:

$$GRG_j = \frac{1}{m} \sum_{i=1}^{m} GRC_{ij} \tag{2.7}$$

2.2. Principal Component Analysis. Principal Component Analysis (PCA) is one of the techniques which has goal to calculate the variation in a sample in as few variables as possible to achieve good accuracy [10]. PCA method summarizes the variation in a correlated multi attribute to a set of uncorrelated components. The advantage of PCA is we can reduce the dimension of the data without losing the information thus the classification accuracy can be maintained. The scheme of the PCA is given as follows

1. Let $X = (x_1, x_2, ..., x_F)$ is the set of features
2. Calculate $\bar{x}_i = \frac{1}{S} \sum_{i=1}^{s} x_i$ as the average of the features
3. Calculate covariance matrix C of the $\bar{x}_i$

$$C = \begin{pmatrix} cov(x_1, x_1 & cov(x_1, x_2 & cov(x_1, x_3 & ... & cov(x_1, x_F \\ cov(x_2, x_1 & cov(x_2, x_2 & cov(x_2, x_3 & ... & cov(x_2, x_F \\ cov(x_3, x_1 & cov(x_3, x_2 & cov(x_3, x_3 & ... & cov(x_3, x_F \\ ... & ... & ... & ... & ... \\ cov(x_F, x_1 & cov(x_F, x_2 & cov(x_F, x_3 & ... & cov(x_F, x_F \end{pmatrix}.$$

4. Calculate the eigenvectors $\mathbf{U} = (\mathbf{u_1}, \mathbf{u_2}, ..., \mathbf{u_F})$ and eigenvalues $\lambda = (\lambda_1, \lambda_2, ..., \lambda_F)$ of the covariance matrix C
5. Calculate the proportion variation of each eigenvalues, where the largest proportion can be assumed containing the largest variance.
6. Compute the principal components scores

$$Y_i = \Sigma_{i=1}^{S} \mathbf{u_i}(x_i - \bar{x}_i) \tag{2.8}$$

7. Compute the correlation between the original data of the features and each principal component.
8. Rank the farthest magnitude values from zero in either positive or negative.
9. We need to determine the level of correlation which will we choose as the dominant features, for example the correlation > 0.5

2.3. Back Propagation Neural Networks. BPNN is a systematic method of training multilayer Back Propagation neural networks. It can perform complex predictions and classification task. It is built on high mathematical foundation and has very good application [15]. BPNN are part of multilayer perceptron (MLP) networks since this model has three layers; an input layer, an output layer and hidden layers which are in between the input and the output layers. The input layer transmits the input data to the hidden layers via the input-hidden weights and bias. Inputs are weighted by the related weights before they are received by the hidden neurons. The hidden neurons process the weighted inputs before sending their outputs to the outputs neuron through hidden-outputs weights and bias. Similar to the inputs data, the output of hidden neuron also weighted by the corresponding weights and processed to produce final output. This structure learns the data input and output connection by repeating the training process until the network produces the correct output. Learning involves incremental changes the connection weights and bias until the network learns to produce the correct output. The final weights are the optimized features of the network. In BPNN, the learning process regarding to obtain the desired results is divided into three phases; these phases are training, validation and testing. Training is a recurrent process to estimate the weights and bias of candidates network design and validation is the recurrent process to estimate the non-training performance error of the candidate network design which is used to stop training once the validation error stops decreasing. Meanwhile, testing is a once-only process after the final weights and bias have been obtained in order to get unbiased estimation on different non-training data. As stated in the explanation above, training is one of the most important phases in determining the pattern of the data input-output. For this research, the training process is focused on classification task in order to obtain good classification accuracy of the bearing conditions. The structure consists of one hidden layers of BPNN, where $x_1, x_2, ..., xn$ are the input variables that make up the input layer. The set of W_{lmn} represents the weights that link the input layer and the hidden layers, where l is the number of iteration of the training process, m and n are the number of neuron in input layer and hidden layer respectively. Meanwhile, b_{l1} and b_{l2} represent the set of bias of input to hidden layer and hidden layer to the output layer. The set of V_{lnp} is the weights which connects the hidden layer and the output layer, where p is the number of neurons in output layer. In the hidden layer neurons, the dot-operation A_n and D_p are conducted, furthermore this additional results pass a nonlinear transfer function $f(A_n)$ to obtain the output O_{1n} from input layer to the first hidden layer and the nonlinear transfer function $f(D_p)$ to obtain the output O_{2p} from the first hidden layer to output layer. There are several transfer functions which are used in BPNN, they are;

Linear	$f(A) = A$
Logistic Sigmoid	$f(A) = \frac{1}{(1+exp(-A))}$
Hyperbolic Tangent Sigmoid	$f(A) = \frac{2}{(1+exp(-2A))} - 1$
Hard Limit	$f(A) = \begin{cases} 1 & \text{if} A \geq 0 \\ 0 & \text{otherwise} \end{cases}$
Gaussian (Radial Basis)	$f(A) = exp(-A^2)$

3. GRA-BPNN and PCA-BPNN. This paper proposed combination techniques of GRA-BPNN and PCA-BPNN to classify the condition of the bearing system. GRA and PCA are applied to determine the dominant features in condition classification. The results from the GRA and PCA are used as the input in BPNN. The scheme of the proposed method is shown in Figure 3.1

The GRA determine the dominant features based on the rank of GRG values while the PCA based on the correlation values between the principal component and the features. The results of these algorithms are presented in the next section.

4. Experimental Results and Analysis. The experiment is conducted by using MATLAB version 7.11 (R2010b) in a computer with Intel Core2 Quad processor Q8200, 2.33 GHz and 1.96 GHz and RAM 3.46 GB. The input of the algorithm are derived from the vibration signal data of the bearings, which are extracted into ten features, namely standard deviation, skewness, kurtosis, the maximum peak value, absolute mean value, root mean square value, crest factor, shape factor, impulse factor and clearance factor [10].The GRA is proceeded to determine the dominant features based on the GRG values thus the dimension of features can be reduced. The GRG values of the features with the distinguishing coefficient $\xi = 0.5$ are given in Table 4.1.

In this paper we choose the five highest GRG values to be identified as the dominant features namely root mean square value, standard deviation, absolute mean value, skewness and maximum peak

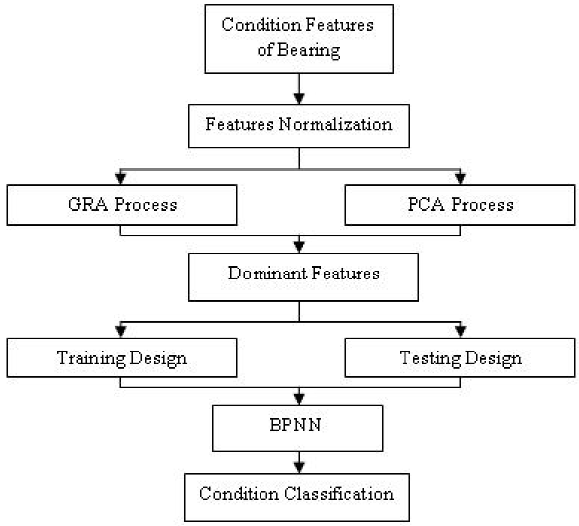

Fig. 3.1. *The Scheme of PCA-BPNN and GRA-BPNN*

Table 4.1
GRG Values of the features

No	GRG	Features
1	0.9448788	Root Mean Sq Value
2	0.9447742	Standard Deviation
3	0.9425986	Abs Mean Value
4	0.9162873	Skewness
5	0.9102795	Max Peak Value
6	0.8334899	Shape Factor
7	0.8297234	Kurtosis
8	0.6041518	Crest Factor
9	0.5358365	Impulse Factor
10	0.4877857	Clearance Factor

value. While for PCA four highest correlation values are chosen as the dominant features, namely root mean square value, standard deviation, absolute mean values and skewness. These dominant features are used as the input in BPNN in order to classify the condition. The test set method is chosen to evaluate the performance of the BPNN structure. Using this method, the 240 data are used which is divided randomly into 80% for training, 10% for validation and 10% for testing respectively. The BPNN structure consists of 30 neurons for input layer, 3 hidden layers with 30 neurons for each hidden layer and 16 neurons for output layer. The BPNN parameters; learning rate and momentum are set as 0.01 and 0.8 respectively. The performance of the algorithms is measured by using the classification accuracy. The classification accuracy is calculated using confusion matrix. Confusion matrix is a method that was proposed to determine the accuracy of the condition diagnosis algorithm. Confusion matrix effectively used for multiclass classification [12]. It provides information about the real data and the experiment result [13] and from that information, it will be easier to obtain the accuracy of true prediction over the total number classes members. Confusion matrix provides the information regarding the actual and predicted output [13]. A confusion matrix with $n \times n$ dimension is associated with actual and prediction of n classification classes.Generally the confusion matrix can be described as

$$\text{Accuracy} = \frac{\text{true prediction}}{\text{total sample}} \times 100\% \tag{4.1}$$

$$\text{Error} = \frac{\text{false prediction}}{\text{total sample}} \times 100\% \tag{4.2}$$

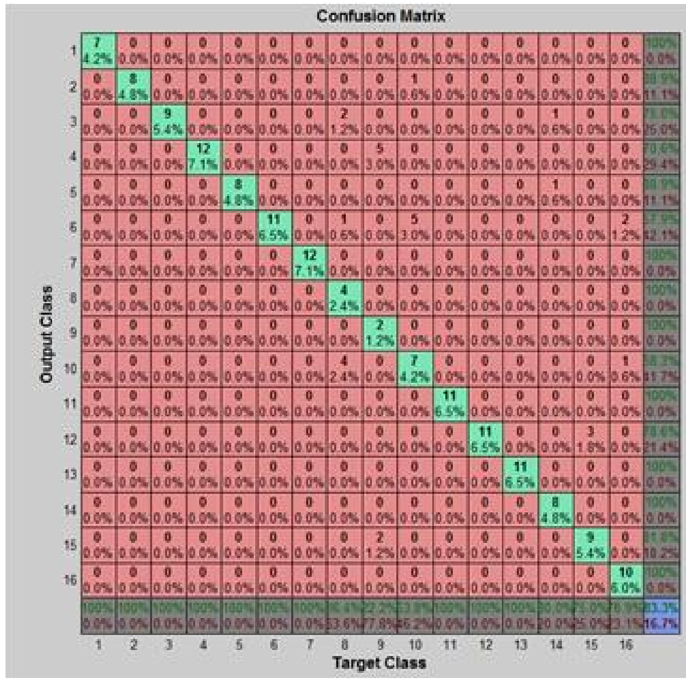

FIG. 4.1. *Confusion Matrix of sixteen class classification*

TABLE 4.2
Classification Accuracy of GRA-BPNN and PCA-BPNN

Algorithm	Iteration	Training (%)	Validation (%)	Testing (%)	CPU Time (s)
GRA-BPNN	60000	82.75	87.58	83.33	2577.48
PCA-BPNN	70000	84.40	80.55	88.31	2776.92

The example of confusion matrix between target class and output class from MATLAB for sixteen class of classification is shown in Figure 4.1. While the classification accuracy of the GRA-BPNN and PCA-BPNN are presented in Table 4.2.

Table 4.2 shows that GRA-BPNN has higher accuracy in validation process, while for training and testing it just achieves 83.33% and 86.11% respectively. Compare with PCA-BPNN accuracy which can achieve 84.4% and 88.31% for training and testing respectively. In CPU time GRA-BPNN needs higher CPU time then PCA-BPNN, for equal number of iteration GRA-BPNN needs 352.24 second longer than PCA-BPNN.

5. Conclusion. This paper proposed two combinations of GRA-BPNN and PCA-BPNN to determine the dominant features of condition classification in bearing system case. The experimental result shows that generally PCA-BPNN has better performance excluding in validation accuracy with four dominant features compared with GRA-BPNN which use five dominant features. PCA-BPNN can achieve higher accuracy in lower processing time compares with GRA-BPNN in the 70000 iterations.

REFERENCES

[1] Cosmin Danut Bocaniala and Vasile Palade. Computational intelligence methodologies in fault diagnosis: review and state of the art. In *Computational Intelligence in Fault Diagnosis*, pages 1–36. Springer, 2006.

[2] Hugo B Fischer, John E List, C Robert Koh, Jorg Imberger, and Norman H Brooks. *Mixing in inland and coastal waters*. Elsevier, 2013.

[3] Avraham Harnoy. *Bearing design in machinery: engineering tribology and lubrication*. CRC Press, 2002.

[4] Tedric A Harris and Michael N Kotzalas. *Essential concepts of bearing technology*. CRC press, 2006.

[5] http://www.dartauto.com/projects/porsche-gt3-wheel-bearing installation/, 2013.

[6] Pratesh Jayaswal, SN Verma, and AK Wadhwani. Application of ann, fuzzy logic and wavelet transform in machine fault diagnosis using vibration signal analysis. *Journal of Quality in Maintenance Engineering*, 16(2):190–213, 2010.

[7] Yiyo Kuo, Taho Yang, and Guan-Wei Huang. The use of grey relational analysis in solving multiple attribute decision-making problems. *Computers & Industrial Engineering*, 55(1):80–93, 2008.

[8] Steve Lawrence, C Lee Giles, and Ah Chung Tsoi. Lessons in neural network training: Overfitting may be harder than expected. In *AAAI/IAAI*, pages 540–545, 1997.

[9] Chun-Hao Li and Ming-Jong Tsai. Multi-objective optimization of laser cutting for flash memory modules with special shapes using grey relational analysis. *Optics & Laser Technology*, 41(5):634–642, 2009.

[10] Weihua Li, Tielin Shi, Guanglan Liao, and Shuzi Yang. Feature extraction and classification of gear faults using principal component analysis. *Journal of Quality in Maintenance Engineering*, 9(2):132–143, 2003.

[11] JL Lin and CL Lin. The use of the orthogonal array with grey relational analysis to optimize the electrical discharge machining process with multiple performance characteristics. *International Journal of Machine Tools and Manufacture*, 42(2):237–244, 2002.

[12] Pierre Machart and Liva Ralaivola. Confusion matrix stability bounds for multiclass classification. *arXiv preprint arXiv:1202.6221*, 2012.

[13] Foster Provost and Ron Kohavi. Guest editors' introduction: On applied research in machine learning. *Machine learning*, 30(2):127–132, 1998.

[14] G. Puscasu, V. Palade, A. Stancu, S. Buduleanu, and G. Nastase. Sisteme de conducere clasice si inteligente a proceselor. *MATRIX ROM, Bucharest, Romania*, 2000.

[15] S. Rajasekaran and G.A. Vijayalakshmi Pai. *Neural networks, fuzzy logic and genetic algorithms: synthesis and applications*. New Delhi, II : Prentice-Hall of India, 2007.

[16] Pedro Vicente Jover Rodríguez and Antero Arkkio. Detection of stator winding fault in induction motor using fuzzy logic. *Applied Soft Computing*, 8(2):1112–1120, 2008.