Performance Evaluation of BPSO & PCA as Feature Reduction Techniques for Bearing Fault Diagnosis

Atik Faysal, Ngui Wai Keng, and M. H. Lim

Abstract Vibration-based signal processing is the most popular and effective approach for fault diagnosis of bearing. In this paper, time-frequency domain analysis, i.e. empirical mode decomposition (EMD) was applied to the raw vibration signal. Intrinsic mode function (IMF) containing the characteristics of vibration data was analysed to obtain 90 statistical features. Two feature reduction algorithms, namely principal components analysis (PCA) and binary particle swarm optimiser (BPSO) were applied individually for feature reduction. The reduced feature subsets were 12 and 35 for PCA and BPSO, respectively. K-Nearest Neighbours (K-NN) was used as an intelligent method for fault diagnosis. K-NN was applied to the entire feature set and individually on the selected feature subset of PCA and BPSO. The reduced feature subset with PCA performed the finest in all the measurements taken. For BPSO, although it effectively reduced the feature dimension and classification time, the testing accuracy was slightly lower. Comparing the output accuracy of the K-NN classifier for the selected methods demonstrated the effectiveness of PCA and BPSO as efficacious feature reduction techniques.

Keywords Feature reduction · Principal components analysis · Binary particle swarm optimiser · Empirical mode decomposition

1 Introduction

Bearing is one of the most used machinery parts of industries that undergoes a huge load of friction, erosion, and different types of failure [1]. It is important to develop a fast and reliable fault diagnosis method to monitor the machine health condition. The incorporation of artificial intelligence with signal processing has been proven to

A. Faysal · N. W. Keng (✉)
Faculty of Mechanical and Automotive Engineering Technology, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia
e-mail: wkngui@ump.edu.my

M. H. Lim
Institute of Noise and Vibration, Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia

A. F. Ab. Nasir et al. (eds.), *Recent Trends in Mechatronics Towards Industry 4.0*, Lecture Notes in Electrical Engineering 730, https://doi.org/10.1007/978-981-33-4597-3_55

be efficient in condition monitoring. Among the different signals, vibration signals have been widely used along with various machine learning classifiers as intelligent diagnosis methods.

Among many signal processing methods, time domain, frequency domain, and time-frequency domain analysis are the most popular ones [2]. The advantage of time-frequency domain analysis over other methods is it provides both time and frequency information. Among the many popular time-frequency domain analysis, short time Fourier transform, wavelet transform, and empirical mode decomposition (EMD) have been widely used. In recent years, a lot of research has been conducted by incorporating EMD with artificial intelligence techniques. Lin et al. [3] combined EMD with FFT to analyse the oil-leakage fault signal of the gearbox of wind turbines. K-NN was applied for fault diagnosis. Pandya et al. [4] processed acoustic emission using EMD and Hilbert–Huang Transform. For classification, a modified K-NN algorithm, where $k = \sqrt{n}$ was proposed based on the asymmetric proximity function. This present study used EMD [5] for feature extraction from vibration signal. By obtaining different statistical features from IMFs, a large number of data can be gathered. In that sense, EMD appears to be suitable for machine learning algorithms. However, a large number of features raise the condition for feature reduction. Two feature reduction approaches, namely PCA and BPSO were applied separately in the current work.

Principle component analysis (PCA) can be a handy tool to perform feature reduction, where the first few columns contain most of the variance of the feature table. PCA is combined with machine learning algorithm to classify different fault classes. Sharma et al. [6] used PCA on the feature of rolling element bearings and the performance was evaluated using weighted K-NN classifier. For three bearing conditions, an accuracy of 100% was achieved using the Mahalanobis distance metric. Among the population-based approaches, the application of particle swarm optimiser (PSO) to features selection has attracted a lot of attention lately. In [7], Binary PSO (BPSO) was implemented with regularised Fisher's criterion, which was used as a fitness function. Later, PCA and BPSO were individually applied to focus on the performance of BPSO. Fei [8] proposed an improved BPSO to select the parameter K of K-NN. The traditional K-NN was trained by varying the parameter K to obtain better accuracy. However, such literature on the comparison between PCA and BPSO using the same classifier is rare. This paper tries to evaluate the performance of PCA and BPSO as effective feature reduction methods. The performance of the reduced feature subset is measured using K-NN classifier. In the end, the performance of the classifiers using the reduced feature subset obtained from the two feature reduction techniques is discussed.

2 Theoretical Background

2.1 PCA

PCA [9] transfers the possible correlated variables into a lower number of uncorrelated variables known as principal components. The first principal component contains the highest possible variability and the following components contain the same and so on. PCA can be an effective tool for dimensionality reduction. It can also be used to reduce the complexity of signal processing. The steps of PCA are as follows:

Step 1: Take the feature set $X_{n\times m}$ where n is the number of data samples and m is the dimension of the feature space.
Step 2: Normalise the data of the entire feature set, $norm(X)$.
Step 3: Take the covariance matrix of the feature set, $cov(X)$.
Step 4: Calculate the eigenvectors and eigenvalues of the covariance matrix.
Step 5: Choose components and form a feature vector.
Step 6: Derive the new data of m-dimension.

2.2 PSO

PSO was proposed by Kennedy and Ebenhart in 1995 [10]. A group of particles (like bird or fish swarm) is assumed and their position and velocity are constantly updated. The velocity and position of the particles are updated according to the following equation:

$$v_{id}^{t+1} = \omega v_{id}^{t} + c_1 r_1 \left(Pbest_i^t - x_{id}^t\right) + c_2 r_2 \left(Gbest_g^t - x_{id}^t\right) \tag{1}$$

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \tag{2}$$

where v_{id}^t denotes the velocity and x_{id}^t is the position of particle after t-th iteration. $Pbest_i^t$ is the particles current best position and $Gbest_g^t$ is the global best position. ω is the inertia weight and d is the dimension of population. c_1, c_2 are the acceleration coefficients. r_1, r_2 are the random numbers uniformly distributed between [0, 1].

2.3 K-Nearest Neighbours

K-NN classifier is a widely used pattern recognition tool. The basic idea of the algorithm is that features belonging to different classes will form clusters in their individual feature space. Neighbours are the features that are close to each other [11].

In the case of a new vector, the K-NN algorithm calculates the distance with the training data using distance matrices, such as Minkowsky, Correlations, Manhattan, Chi-square, and Euclidean distance. From the result, the nearest neighbours can be calculated and the output class can be obtained. A class is assigned to the new vector from the class pool of its k nearest neighbours, here, k represents the number of the nearest points considered.

3 Data Collection

In this study, for fault diagnosis, the vibration data were obtained from the online dataset of the Case Western Reserve University Bearing Data Center [12]. Acceleration data were measured from a 2 hp Reliance Electric motor bearings. The collected vibration signals included the following operating conditions: (1) normal condition/baseline condition, (2) inner race fault, (3) outer race fault, and (4) ball fault. Each fault condition included three different fault sizes, i.e. 0.007, 0014, and 0.021 in.

4 Feature Extraction

For feature extraction, 1202 data points were taken per sample and the number of total samples was 900. The number of samples was 90 for the baseline condition and each fault condition had 270 samples individually. The time-frequency domain analysis, i.e. EMD simply decomposes a vibration signal into several IMFs. The IMF function must meet two basic conditions:

(a) The number of extrema and zero crossings in the whole vibration series must be the same or differ at most by one.
(b) The mean of upper and lower envelops must be zero at any given point.

It is observed that for the later values of IMFs, the strength of the signal decreases. Therefore, in this experiment, only the first five IMFs were taken for feature extraction from every observation. Different statistical and time-domain features were extracted from the obtained IMFs of drive end and fan end vibration data. From each IMF, nine different features were extracted, which were mean, root mean square, variance, standard deviation, kurtosis, skewness, peak, crest factor, and energy. From the five IMFs, 45 features were extracted in total from the drive end and fan end individually. Juxtaposing all the features, the size of the feature vector became 900×90 for all number of observations. All features were normalised between (0, 1) to ensure the elimination of overfitting while classification.

5 Feature Reduction

The output accuracy of the classifier algorithm depends largely on the features of the dataset. However, all of the features do not have an equal contribution to the performance of the classifier. There might be some features that are not as important as others. Moreover, for a large number of features, the feature set may become redundant and overlapping may occur between different classes. This may result in a reduction of the performance of the classifier with longer computational time. For this reason, the elimination of insignificant features is important. Two different feature reduction methods were used in this study, namely PCA and BPSO. These two methods will be discussed elaborately in the following subsection.

5.1 Feature Reduction Using PCA

PCA was applied to the raw feature set and the values of the score and explained variance were obtained. The score represents an equal number of new variables in a compact feature space and the explained variance is the weight of importance [13]. From the percentage of the explained variance, the optimal number of principal components was selected. Later, by increasing the number of the principal components, K-NN was applied to the selected feature subset and each output was recorded. The threshold level of accuracy for the minimal number of principal components was obtained thereby.

5.2 Feature Reduction Using BPSO

The proposed BPSO model, which combines K-NN for its fitness function evaluation was used to select the best features from the feature set to improve the classification ability. In BPSO, the number of initial population is taken the same as the number of features, where the positions of features are initialised with random 0's and 1's. For example, X = [0 1 1 0 1 0 0 1 … 1] is a position vector, where bit 1 denotes that the feature is selected and bit 0 denotes that the feature is not selected. Therefore, new feature subsets are generated and for a population of N_p features, N_p feature subsets is generated. The ultimate goal is to find the best feature subset that maximises the class distinguishability.

The training error of K-NN classifier was used as the fitness function of the BPSO algorithm and the model was trained to minimise this fitness function. The parameter setting for K-NN was such that the number of neighbours was two and 2-fold cross-validation (CV) was used for data partitioning. The 2-fold CV divided the training sample into two equal subsets, among which one was used for training the model.

The training accuracy of K-NN, A_i, was obtained and by using that the fitness of ith particle was defined as follows:

$$fit_i = 1 - A_i \tag{3}$$

The training error was evaluated in each iteration and the personal best, $Pbest$ and the global best, $Gbest$ were updated.

In the present work, to select the most desired features from the features pool (90), BPSO was implemented. To measure the accuracy of the selected feature subset, K-NN was applied. The BPSO was initialised with the following parameters:

- The initial population is the same as the particle size, i.e. 90 particles, where particle size is the number of all features (as recommended by [14])
- $\omega_{\text{min}} = 0.4$, $\omega_{\text{max}} = 0.9$, $v_{min} = -2$, $v_{\text{max}} = 2$, $c_1 = 2$, $c_2 = 2$, r_1 and r_2 are randomly generated between 0 and 1.
- Number of iterations, $N_i = 200$.

5.3 K-NN Modelling

In this study, the Euclidean distance metric was considered for the K-NN model and the number of neighbours was two [15]. To avoid overfitting due to the large number of samples (900), 50% holdout cross-validation (CV) [16] was used, meaning half of the dataset was used for training the classifier and the other half was used for testing. The computational platform used was MATLAB 2017b and the computation unit used was Intel core i5, CPU 2.3 GHz, 4 GB RAM.

6 Results and Discussion

6.1 Implementation of K-NN on the Raw Feature Set Without Feature Reduction

Firstly, K-NN was applied to the entire feature set without passing through any feature reduction methods. The objective was to observe the performance of the classifier without having the feature dimension reduced and later, this result was used to compare the performance of the selected feature reduction methods. From the output of K-NN, the inner race fault and outer race fault conditions were completely successful to be classified with around 5% error for both the baseline and ball fault conditions. The overall testing accuracy of the classifier was 97.78%.

6.2 Implementation of PCA

PCA was applied to the entire feature set for feature reduction. Figure 3 was obtained from the cumulative summation of the explained variance for the increasing number of principal components. This helped to find the minimum number of the principal components that should be considered for classification.

Based on Fig. 1, it can be observed that at 20 principal components, more than 95% of the feature set is covered. At 40, the variance is over 99% and at around 70, the variance level reaches 100%.

Next, the K-NN was applied to the first 40 principal components because it covered almost all the variance. The effectiveness of these components on the classifier performance was determined. The accuracy of every single component was stored and the results in Fig. 2 were obtained. It was evident that there was a sharp increase in accuracy until the first four principal components. The accuracy was 98% at only four principal components and it reached 99.56% at 12 principal components. After this point, there was not much change in the accuracy except for a slight variation. Therefore, it is safe to assume that at 12 principal components, the classifier provides the best accuracy.

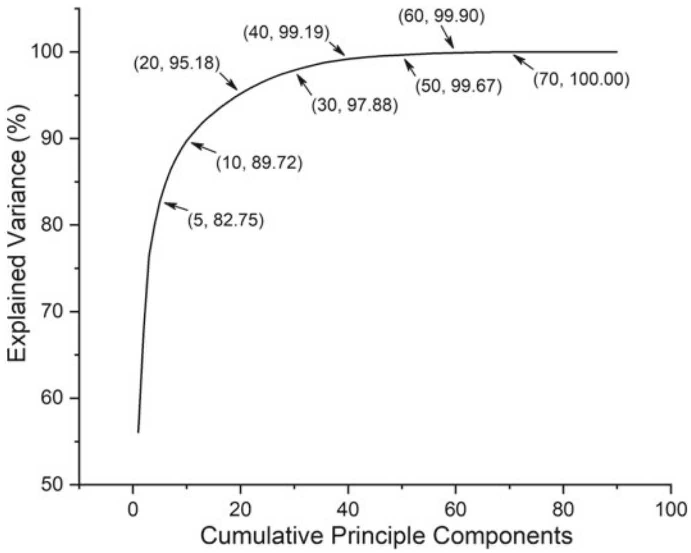


Fig. 1 The rate of explained variance of individual principal components

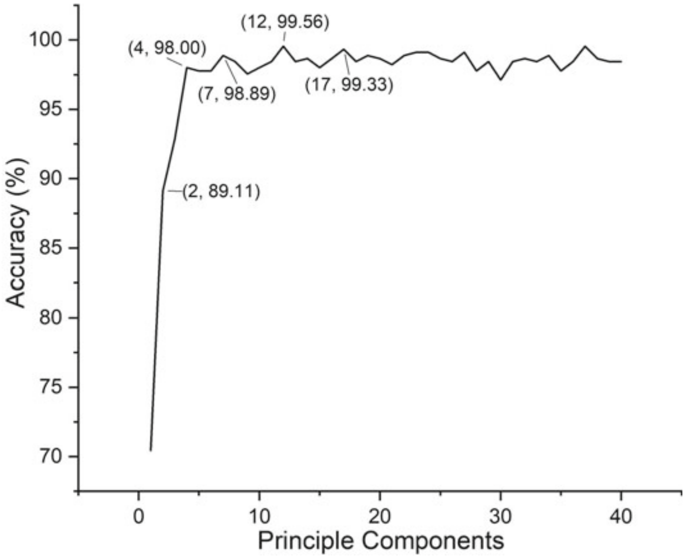
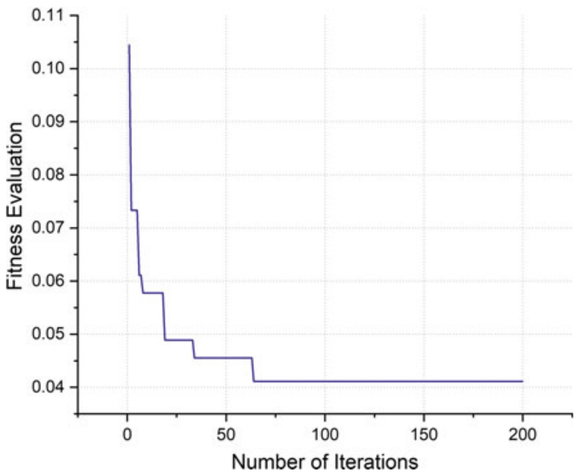


Fig. 2 Accuracy of K-NN using the different numbers of principal components

Fig. 3 Convergence curve for the fitness function of BPSO



6.3 Implementation of BPSO

By looking at the convergence curve of the proposed BPSO + K-NN based features selection algorithm, the result can be analysed. In Fig. 3, the Y-axis and the X-axis represents the number of iterations. It shows that around 55 generations, the algorithm reaches the global best solution. This proves that an adequate number of iterations is provided.

Table 1 Comparison of test accuracy and classification time of different methods

Method	No. of features	Accuracy (%)		Classification time (s)
		Training	Testing	
K-NN	90	99.33	97.78	0.111
PCA & K-NN	12	100	99.56	0.076
BPSO & K-NN	35	99.78	97.11	0.086

After applying BPSO, the number of selected features was 35, which implied that among the 90 initial features, the selected ones were the most efficient. These selected features were used for classification using K-NN, where the parameters of K-NN were the same for the entire feature set. The results will be discussed in the next section.

In this part, the output accuracy of the K-NN classifier was listed with the reduced feature subset of PCA and BPSO for comparison. In summary, three different methods were applied in the present work, i.e. (1) classification with no feature reduction, (2) classification with PCA, and (3) classification with BPSO. The following table depicts the accuracy and computation time of the methods.

Based on Table 1, it can be observed that the K-NN method, while taking all the features, requires the longest computation time. For PCA, with only 12 principal components, the highest accuracy of training and testing is obtained. The training accuracy is 100% for PCA, whereas the testing accuracy is 1.78% higher than the one with no feature reduction. Therefore, the number of dimensions is reduced from 90 to only 12 with less computation time using PCA. This result suggests that the feature subset obtained by PCA is more effective for classification than using the entire feature set.

For BPSO, the 35 features subset provided accuracy close to the accuracy of that of the entire features with a significant reduction of computation time. The testing accuracy of the BPSO was only 0.67% less than the performance of the K-NN with no feature reduction. Not only the features were significantly reduced but also it provided better training accuracy. The possible explanation for the reduction of testing accuracy is, unlike PCA, BPSO does not transfer the features by projecting them onto a set of orthogonal axes. Instead, it chooses the most dominant features using a metaheuristic algorithm, which includes the stochastic elements. Although being the least significant, the unselected features carry some information that contributes to the classification. However, for its significant reduction in dimension and computational time, a slight reduction of accuracy might be considered.

7 Conclusion

In this paper, IMFs from EMD were used to obtain different statistical features. The IMFs provide many details of vibration signals, thus producing sufficient features.

The proposed two feature reduction methods, i.e. PCA and BPSO proved to be effective in terms of reducing features, improving accuracy, and deducting computational time. The threshold accuracy of the K-NN classifier was obtained using only 12 principal components from 90 features. PCA seemed to do well comparatively in terms of the number of reduced features and classification accuracy. The number of selected features using BPSO was 35, which proved to be adequate for classification. Although BPSO successfully reduced the number of features and classification time, its testing accuracy was slightly less. This leaves the scope for improving BPSO by modifying its parameters. In the future, the authors wish to use different fitness functions with more suitable initialisation parameters. The results in this study justified the importance of PCA and BPSO as effective feature reduction methods.

Acknowledgements The authors are grateful to the Government of Malaysia for the financial support provided under the FRGS/1/2018/TK03/UMP/02/24 grant. Additional funding for this research also came from Universiti Malaysia Pahang (PGRS200303) and the Institute of Noise and Vibration, Universiti Teknologi Malaysia (RDU192303).

References

1. Liu Z, Zhang L (2020) A review of failure modes, condition monitoring and fault diagnosis methods for large-scale wind turbine bearings. Meas J Int Meas Confed 149:107002
2. Wei Y, Li Y, Xu M, Huang W (2019) A review of early fault diagnosis approaches and their applications in rotating machinery. Entropy 21(4). https://doi.org/10.3390/e21040409
3. Lin DF, Chen PH (2012) Fault recognition of wind turbine using EMD analysis and FFT classification. In: Proceedings of 2012 3rd international conference on digital manufacturing & automation (ICDMA 2012), pp 414–417
4. Pandya DH, Upadhyay SH, Harsha SP (2013) Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN. Expert Syst Appl 40(10):4137–4145
5. Cheng Y, Wang Z, Chen B, Zhang W, Huang G (2019) An improved complementary ensemble empirical mode decomposition with adaptive noise and its application to rolling element bearing fault diagnosis. ISA Trans 91:218–234
6. Vernekar K, Kumar H, Gangadharan KV (2017) Engine gearbox fault diagnosis using empirical mode decomposition method and Naïve Bayes algorithm. Sadhana Acad Proc Eng Sci 42(7):1143–1153
7. Ziani R, Felkaoui A, Zegadi R (2017) Bearing fault diagnosis using multiclass support vector machines with binary particle swarm optimization and regularized Fisher's criterion. J Intell Manuf 28(2):405–417
8. Fei SW (2019) The HYBRID METHOD of VMD-PSR-SVD and improved binary PSO-KNN for fault diagnosis of bearing. Shock Vib 2019:1–8
9. Nimityongskul S, Kammer DC (2010) Frequency domain model reduction based on principal component analysis. Mech Syst Signal Process 24(1):41–51
10. Kennedy J, Eberhart R (1995) Particle swarm optimization—IEEE conference publication. In Proceedings of ICNN'95—international conference on neural networks, vol 4. IEEE, pp 1942–1948
11. Rashid M, Sulaiman N, Mustafa M, Bari BS, Sadeque MG, Hasan MJ (2020) Wink based facial expression classification using machine learning approach. SN Appl Sci 2(2):1–9

12. Loparo KA (2013) Bearing data center. Case Western Reserve University. https://csegroups.case.edu/bearingdatacenter/pages/download-data-file
13. Shin EC, Craft BD, Pegg RB, Phillips RD, Eitenmiller RR (2010) Chemometric approach to fatty acid profiles in Runner-type peanut cultivars by principal component analysis (PCA). Food Chem 119(3):1262–1270
14. Samanta B, Nataraj C (2009) Use of particle swarm optimization for machinery fault detection. Eng Appl Artif Intell 22(2):308–316
15. Sharma A, Jigyasu R, Mathew L, Chatterji S (2018) Bearing fault diagnosis using weighted K-nearest neighbor. In 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, pp 1132–1137
16. Křížek P, Kittler J, Hlaváč V (2007) Improving stability of feature selection methods. In Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), vol 4673. LNCS, pp 929–936