Using PCA in Acoustic Emission Condition Monitoring to Detect Faults in an Automobile Engine

Peyman KABIRI 1, Amir MAKINEZHAD 2

1 Intelligent Automation Laboratory, School of Computer Engineering, Iran University of Science and Technology, 16846-13114, Tehran, Iran, Tel: +98 (21) 77240540 to 50, Ext. 3341; Fax1: +98 (21) 73223341, peyman.kabiri@iust.ac.ir
2 Intelligent Automation Laboratory, School of Computer Engineering, Iran University of Science and Technology, 16846-13114, Tehran, Iran, a_makinezhad@comp.iust.ac.ir

Keywords: Acoustic Emission, Fault Diagnosis, Wavelet, PCA

Abstract

Automobile industry is an important industry in many countries. Monitoring the operating conditions in automobiles is a big challenge not only when the product leaves the factory but also during the after sale maintenance programs. This paper proposes a new approach in acoustic emission based condition monitoring and fault detection that uses wavelet method. The fault that was investigated in this study is related to ignition system of the automobile, more specifically speaking, operation of the sparks. The sampled dataset includes sound samples collected in repair shops. Noise was a major problem for this work. Recording the sound signals in repair shop will include unwanted noise. Human voice, sound of the wind and the sound made by other operating automobile engines are also recorded in the sampled sound record, where, some of them may include arbitrary and random noise. Energy, RMS, Kurtosis, Skewness, Marse, Crest Factor, Zero count Minimum count and Maximum count features are calculated from recorded sound signals. Principal Component Analysis (PCA) is used for dimension reduction. These new orthogonal features are used for classifying recorded data. As a proof for generalization capability of the method, 10% of the dataset was uniformly selected for the training dataset and the remaining 90% are used for the test. 150 sampled records from the acoustic emissions out of similar models of a car (Kia Motors under license products) were recorded; one sample record for the healthy engine and one sample for the faulty engine (300 sampled records in total). Accuracy of the proposed method is above 70% percent. Article describes the laboratory tests which make the first stage of the study concerning the use of the AE method to determine the technical state of the slide bearings in engines with self-ignition.

Introduction

In the recent years, automobile industry has grown rapidly. Maintenance of the products is one of the important parts of automobile industry. Cost of condition monitoring and fault detection is high. Inexpensive methods for condition monitoring and fault detection are useful to reduce the maintenance costs, e.g. evaluation of the operating condition of an automobile engine. Acoustic Emission Testing (AET) is a type of non-destructive testing methods and is widely used in the industry. For example, it is used for condition monitoring and fault detection in components such as bearings, gearboxes, engines and rotating structures [1]. Ghiurcau and Rusu [2] used Tespar alphabet and Tespar matrix to recognize type of the vehicle. In another reported work, Wu and Liao [3] used wavelet analysis to diagnose fault in automobile air-condition blower by acoustic emission signals. Sound of the automobile engine is a stochastic signal [4]. Ghiurcau and Rusu in their work, assume that if automobiles in the same class have the same operating condition then the acoustic emission of their engine will be the same [2]. Problem addressed in this work is to find the fault in automobile engine using its acoustic emissions. Fault diagnosis for the engines
has been addressed in many research reports. However, in almost all of the reported works, researchers have implemented and tested their methods inside the laboratory, where, environmental noise is not considered.

“Acoustic emissions (AEs) are defined as transient elastic waves generated from a rapid release of strain energy caused by a deformation or damage within or on the surface of a material” [1]. AE is used in different fields of work for condition monitoring and fault diagnosis. AE signals of the automobile engine are stochastic and non-stationary [5] and within a wide range of frequencies i.e. about 10Hz to 2MHz [6]. Some researchers have used high frequency sensors for fault diagnosis and condition monitoring [7] [8]. In other works, low frequency sensors such as microphones are used [9]. Find a specific feature for fault in automobile engine by AE is very difficult [10]. Wu et al. [3] used frequency domain features and image processing algorithm to detect fault.

Aim of the reported work in this paper is to derive effective features for fault diagnosis in AE of an automobile engine. In the purposed method of approach, a large number of features are initially calculated, later on, using PCA method effective features are selected. Fig 1 depicts flow diagram used to derive desired features in the proposed method.

**Features**

AE characteristics are mentioned in many articles. Each characteristic provides a series of specific information about the studied AE signal. Energy, RMS, kurtosis, skewness, Marse, crest factor, zero count, minimum count and maximum count features are used in this study. The aforementioned time domain features are explained in the following:

**Energy**

Let \( x(t) \) is the AE signal. The Energy of the signal in discrete mode is defined by the EQ 3 [8].

\[
E = \sum_{i=a}^{b} |x(t)|^2
\]  

**RMS**

Root Mean Square (RMS) is sum of squares difference between the amplitudes and mean value of the signal within a specific period (EQ 4) [1].

![Fig.1. Flow Diagram of proposed method](image-url)
Kurtosis
Kurtosis measures the density of points around the mean value. In other words, kurtosis is a criterion for peakedness of the signal (EQ 5) [11].

\[
\gamma_4 = \frac{\sum (x_i - \bar{x})^4}{(n-1)\sigma^4} - 3
\]  

(5)

Skewness
Skewness is a criterion for measuring the asymmetry in the signal. If the value of skewness is negative, most of the data are scattered on the left (or top) side of the average line and vice versa. The skewness is zero if the statistical data distribution is normal (EQ 6) [11].

\[
\gamma_1 = \frac{\sum (x_i - \bar{x})^3}{(n-1)\sigma^3}
\]  

(6)

Marse
Marse is a feature that measures the surface under the curve. Marse is calculated for positive and negative areas separately [1].

Crest factor
It is the difference between absolute maximum and RMS values [1].

Counts
Number of crossings from horizontal axis is zero crossing counts and the number of local maximums and local minimums of the signal within time frame are used as features in this approach [1].

Cumulative counts
Number of crossings from different threshold values is represented by different cumulative counts [8]

The following parameters present time-frequency domain features:

Wavelet Transform
Wavelet transforms are inner product of signal and a family of the wavelets. The family of the wavelets can be calculated \( \psi(x) \) as the mother wavelet (EQ 7) [12].

\[
\psi(a,b) = |a|^{1/2} \psi \left( \frac{t-b}{a} \right)
\]  

(7)

Where \( a \) is scale factor and \( b \) is time location (EQ 8).
\[
    w(a, b) = |a|^{1/2} \int x(t) \psi_{a,b}^* dt
\]

Support Vector Machine (SVM) is used in this study for classification and calculation accuracy [13].

Experiments

The sampled dataset is collected in a repair shop. PRIDE automobile (KIA motors) is selected for the fault diagnosis [14]. The selected fault for investigation is within the ignition system. The fault scenario is to have the engine work with combustion occurred in only 3 out of 4 cylinders. To do so, spark connection for a certain cylinder is disconnected.

Dataset was collected in MP3 format using Kingston voice recorder. To do so, the hood of the car was opened and the recorder was placed at the certain position above the engine. Recording time for each sampled sound is 60 seconds with 128Kbps bit rate and audio sample rate of 32 KHz in stereo. Half of the recording time belongs to the sound of the engine operating in healthy condition, and the rest is for the sound of the engine operating with fault.

Pre-processing of the sampled data

The recorded sampled data was manually analysed to separate samples with extremely high noise and to remove them. In this cleansing process some parts of the signal is deleted. Removed parts include human voice and sound of the other automobiles working close to the test subject whom have high volume. The resulted dataset consists of the sound of engine operating in healthy and faulty conditions with the recording time of 15 seconds each.

Test and train datasets

The sampled dataset was collected from 150 Pride automobiles and the dataset consists of two classes. Class A has 150 sound samples of healthy engines and class B has 150 sound samples of faulty engines. The total number of dataset is 300 sound samples.

The total dataset was divided into a training dataset and a test dataset. In order to prove the generalization capability of the proposed method, 10% of the original dataset was selected as the training dataset and 90% as the test dataset. Obviously, once the number of samples in the training set increases the accuracy of the result will improve.

Feature extraction

Features are extracted in both the time and the time-frequency domains. The sound signal has to be windowed since the automotive sounds are non-stationary. Assuming small size for the time window, signal can be considered to be stationary. Applying the aforementioned window on the signal features such as energy, RMS, skewness, kurtosis, crest factor, Marse, zero counts, maximum counts, minimum counts and cumulative counts with 100 different thresholds are extracted. In all the above time windows, eight new derived features are calculated from a combination of the features extracted in previous stage. The features extraction is explained in fig 2. Consequently, 880 features are derived from each sample.
Once time domain features are calculated, Meyer Discrete Wavelet Transform (Meyer DWT) is used to calculate the time-frequency domain features. Detail coefficients and approximation coefficients are calculated using Meyer DWT. Mean, variance, skewness, kurtosis and energy of these coefficients are also calculated. In other words, 10 additional features are calculated using Meyer DWT. Fig 3 presents the normalized time domain and wavelet features matrix.

Without any additional pre-processing, using these features for classification will result in 64.58% of accuracy for the fault detection. Results are presented in table 1. This accuracy is not acceptable and training the classifier takes a long time. Improving the performance of the method, PCA was used to reduce the dimensionality of the feature space.

![Fig. 3. Normalized features extracted from AE signals](image)

Table 1- Classification results with original number of features.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Positive Rate</td>
<td>61.51%</td>
</tr>
<tr>
<td>False Positive Rate</td>
<td>32.36%</td>
</tr>
<tr>
<td>True Negative Rate</td>
<td>68.39%</td>
</tr>
<tr>
<td>False Negative Rate</td>
<td>38.49%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>64.71%</td>
</tr>
</tbody>
</table>
Principal Component Analysis

Principal Component Analysis (PCA) is a method used for dimension reduction and feature selection. Calculating PCA matrix, covariance matrix must be calculated. Fig 4-left shows the covariance matrix of the sampled dataset, and corresponding PCA matrix is presented in fig 4-right.

The first 10 Principal Components (PCs) hold the largest variances the derived variables before a sharp decline in the variance of the PCs.

The sorted PCA coefficient matrix is presented in fig 5-right. Fig 5-right explains the importance of the individual features in the production of variables. Classification results are reported in table 2.

Training Time 0.0058 seconds

Fig. 4. Left: Covariance matrix for feature space – Right: PCA coefficients for the features sorted with respect to their corresponding eigenvalues.

Fig.5. Left: Eigen values of the eigen matrix – Right: Sorted PCA coefficients matrix
Table 2- Classification results after dimension reduction.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>True Positive Rate</td>
<td>73.56%</td>
</tr>
<tr>
<td>False Positive Rate</td>
<td>31.11%</td>
</tr>
<tr>
<td>True Negative Rate</td>
<td>71.11%</td>
</tr>
<tr>
<td>False Negative Rate</td>
<td>24.44%</td>
</tr>
<tr>
<td>Accuracy</td>
<td>72.64%</td>
</tr>
<tr>
<td>Training time</td>
<td>0.00045 seconds</td>
</tr>
</tbody>
</table>

As reported in table 1 and 2, false positive rates are still high. This problem will be addressed in future work. Those features that are selected by the PCA mainly include frequency time domain features derived from the wavelet. Comparing the training time from the first and second experiments (tables 1 & 2) it can be seen that the training time using the reduced feature space is 12.89 times shorter than the training with original feature space. This is true while the true positive rate and accuracy for the results derived from the reduced feature space is also improved.

**Conclusions**

In this study a new method for finding fault signature is proposed. Intention is to find effective features for fault classification. In this work extreme noise levels are eliminated, however, environmental noise is still present within the dataset.

Using only 10% of the dataset for the training and 90% for the test provides the insurance about the generalization capability of the proposed method. Therefore, it is expected to be able to apply this method on different types of vehicles. Having wavelet based features selected by PCA is an indication of the suitability of these features for the purpose of the AE classification. This is due to the fact that wavelet is robust with respect to the noise. Consequently, using the reduced feature space, positive rate and the training speed are improved.

**Future works**

As for the feature work, intention is to improve the quality of the results of the classification by selecting a more suitable feature set classification methods. Therefore, plan is to put more emphasis on the time-frequency domain features. It is also expected to make the sound dataset accessible to the public via our laboratory website at (http://ial.iust.ac.ir/).

**References**