Vibration signatures, wavelets and principal components analysis in diesel engine diagnostics

G. Chandroth, A J C. Sharkey & N E. Sharkey
Department of Computer Science, University of Sheffield, 211 Portobello Street, Sheffield S1 4DP, UK
EMail: gopi@dcs.shef.ac.uk

Abstract

The vibration signatures of a normally aspirated diesel engine contain valuable information on the health of the combustion chamber components. It could be used to detect incipient faults in the engine. Several commonly occurring faults were induced in a 4-stroke diesel engine and the ensuing vibration signals recorded. Three different feature extraction techniques: Domain expertise, wavelet analysis and principal components analysis (PCA), were compared to evaluate the effectiveness of each method. The extracted features were then used to develop artificial neural nets based fault diagnostic systems. It was found that the best results were obtained with the wavelets based feature extraction technique. However, each of the three systems was shown to exhibit a high degree of accuracy. In a separate study, ensembles of 3 nets were created. The majority of ensembles performed better than any of the constituent nets, thus demonstrating the power of the technique of combining neural nets in majority voting systems.

1 Introduction

Vibrations in a reciprocating internal combustion engine are caused by unidirectional combustion forces and mechanically induced structural resonances (Preide[4]). In the vibration waveform sensed from a cylinder head, the predominant components in the signal are due to the combustion forces. It has been shown by Preide[4] that the gas pressure oscillations during combustion have a very strong correlation to the vibration patterns. The various sub–assemblies and components participating in the engine combustion, such as air inlet and exhaust valves, fuel injectors, pistons, piston rings and cylinder liners, have very specific functions associated with them. It is expected that any anomaly in these components would cause deviations in the normal gas pressure oscillations and the consequent vibration signatures would have strong features relating to the particular anomaly. We validate this approach by developing neural nets based fault diagnosis systems trained on vibration data acquired during normal operation (N) and after inducing four commonly occurring faults in the engine components (poorly atomising fuel injector I, blocked injector nozzle B, leaking air inlet valve L and leaking exhaust valve E). The severity of each of the faults induced was very low.

An investigation of the validity of the approach of training neural nets to identify known faults on the basis of vibration data requires the application of a preprocessing method to reduce the dimensionality of the input. The original signal consisted of 7200 points representing a complete engine cycle. The effectiveness of three pre-processing methods: wavelets, PCA and domain expertise, were compared. In section **??** we introduce the theory behind wavelet transforms followed by the application of orthogonal wavelet analysis to the engine vibration data. We illustrate a novel method for choosing the best discriminant features from each level of wavelet decomposition. In section **??** the same vibration data is analysed using principal components analysis, projecting the high dimensional data onto orthogonal basis of low dimension. In section **??** we utilise domain knowledge of the combustion characteristics of a diesel engine, to pick a small subset of the original data sample.

In each of the sections mentioned above, we demonstrate how the features extracted could be used to train artificial neural nets capable of classifying the 5 classes to a high degree of accuracy. Finally, in section **??** we utilise the various nets developed as mentioned above, and combine them into several majority voting systems most of which are found to exhibit better generalisation capability than any of their individual constituents.

2 Wavelet transforms

Wavelets (small waves) are a set of basis functions $W_{ab}(t)$in continuous time, constructed from a single mother wavelet $W(t)$. The wavelet transform decomposes an arbitrary signal $x(t)$ into a superposition of elementary functions obtained by a process of dilating and translating the mother wavelet. The wavelet coefficients obtained from this process can be considered as similarity indices between the section of the compared signal and the analysing wavelet. The continuous wavelet transform is defined as (Chui[3])

$$W_{\psi}^{x}(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\, \psi^{*}\left(\frac{t-b}{a}\right) dt, \tag{1}$$

where $\psi(t)$ is the wavelet function, b ($b \in \Re$) is a translation parameter indicating the locality, a ($a > 0$) is a dilation or scale parameter and $*$ denotes complex conjugation. Continuous wavelet transforms create redundant information. A discrete set of translation and dilation parameters are often sufficient for most tasks. Wavelet transform within the dyadic framework, i.e. $a_j = 2^j$ and $b_{j,k} = k/2^j$, defines the orthogonal wavelet transform. A function $\psi(t)$ is called an orthogonal wavelet if the family

$$\psi_{m,k}(t) = 2^{m/2}\psi(2^m t - k), \tag{2}$$

where $m, k \in Z$, forms an orthonormal basis of functions. The orthogonal wavelet transform then becomes

$$x_k^m = \int_{-\infty}^{\infty} x(t)\psi_{m,k}(t)dt, \tag{3}$$

The orthogonal wavelet transform represents the most efficient, yet parsimonious representation of the original signal. The coefficients of the transform could be used as input features for pattern recognition systems. In analysing the engine vibrations we use Daubechies wavelets (Daubechies[4]) of the 8^{th} order (DB_8) which has compact support (the function has zero value outside a certain interval or support) and 8 vanishing moments (the degree of the polynomial upto which the wavelet coefficients remain small). In the present task (feature extraction), the higher number of vanishing moments and orthogonality are desirable properties as we are interested in the non redundant high frequency components of the signal.

2.1 Decomposition of vibration data

Wavelet coefficient variance has been used by Staszweski et al[9] as inputs to pattern classifiers. Here we use the covariance between coefficients from the 5 classes to identify a subset of coefficients for training neural nets. There were 12000 data samples in total corresponding to the 5 classes (N,I,B,L and E), in equal proportion of 2400 samples per class. Each data sample was a vector of length 7200 points representing one complete engine cycle (2 revolutions of the 4-stroke engine). From each sample, a section of 1800 points, representing the data from bottom dead center (BDC) up to the top dead center (TDC) of the combustion zone, was selected. Each set of data belonging to the 5 classes were then decomposed to 8 levels using the DB_8 orthogonal wavelet transform. The decomposition process consists of convolving the analysed signal with low pass and high pass filters made up of Daubechies filter coefficients of the 8th order. Applying a recursive process of filtration and down-sampling the signal length is reduced to a point beyond which it cannot be decomposed. The algorithm allows a maximum of $log_2 N$ levels of decomposition, where N is the number of data points in the analysed signal.

After decomposing to 8 levels, the potential pool of data sets increased eight fold to 96000 samples consisting of wavelet coefficients at each level. Eight separate matrices M_1 to M_8 were then assembled from this pool. Each M_i consisted of 2400 rows and a varying number of columns decreasing approximately by half from level to level (907, 461, 238, 126, 70, 42, 28, 21 for M_1 to M_8). The elements in the matrices were the wavelet coefficients at each level representing each of the 5 classes.

In order to extract the coefficients with the maximum discriminant information across the five classes, the covariance matrices $\mathbf{\Sigma_k}$ for each M_k $k = 1...8$ were computed. The ij th element of $\mathbf{\Sigma_k}$ represents the covariance between column i and column j. Thus the diagonal of $\mathbf{\Sigma_k}$ is an indication of the point by point variance between the data samples of the 5 classes. Figure **??** is the plot of the diagonal of each $\mathbf{\Sigma_k}$.

Using Figure **??** as a guide, a subset of the points from the rows of the matrices M_k were selected, for developing neural nets based fault diagnostic systems (refer Table **??**). Each data set was then split into 3 parts, retaining 7500 samples for training, 1500 for validating and 3000 for testing. 1 of k output class neural net models with

Level	Total coeff.	Coeff. used	Perf. %
1	907	40	94.07
2	461	40	90.00
3	238	40	93.73
4	126	30	95.73
5	70	15	97.40
6	42	9	79.47
7	28	18	91.20
8	21	14	83.87

Table 1: Performance of various neural nets based on a subset of the wavelet coefficients at 8 different levels.

the backpropagation algorithm were then trained to classify the 5 classes. The number of coefficients selected from each level and the generalisation performance of the neural nets based on coefficients from each level are shown in Table **??**.

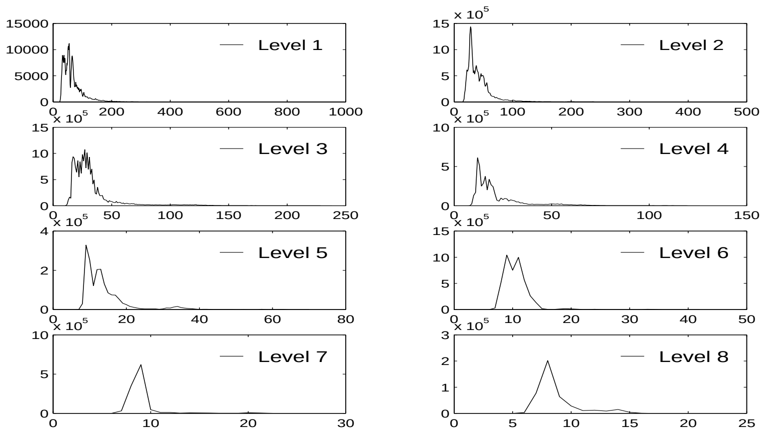

Figure 1: Diagonal of the covariance matrices constructed from 8 different levels of the DB_8 wavelet transform of the five classes of data (N, I, B, L and E)

3 Principal components analysis

Principal components analysis (PCA) is a linear technique which transforms an m dimensional vector $\mathbf{x_i}$ into an n dimensional vector $\mathbf{z_i}$ where $n < m$. This transformation is carried out by projecting the original vector $\mathbf{x_i}$ onto new orthogonal vectors that span a lower dimensional space, creating a new vector $\mathbf{z_i}$. It can be shown (Bishop[1]) that the minimum of the sum of the squared errors between all $\mathbf{x_i}$ and all $\mathbf{z_i}$ is reached when the basis vectors $\mathbf{u_i}$ meet the condition

$$\mathbf{\Sigma u_i} = \lambda \mathbf{u_i} \tag{4}$$

where $\mathbf{\Sigma}$ is the covariance matrix of $\mathbf{x}$, λ is the vector of eigenvalues and $\mathbf{u}$ is the eigenvector matrix of $\mathbf{\Sigma}$ corresponding to λ. Thus, choosing the first few columns of $\mathbf{u}$, based on the first l maximum eigenvalues and projecting any new $\mathbf{x_i}$ on to this reduced space gives us the basis transformed and dimensionally reduced vector $\mathbf{z_i}$. For a detailed discussion on PCA, see Joliffe[5].

3.1 PCA on vibration data

From each original vibration data vector representing an engine cycle, 200 points in the range 180^o to 200^o (combustion zone up-to TDC), were selected. It was then decided to further reduce the data dimensionality using PCA. A plot of the eigenvalues obtained by the analysis is shown in Figure **??**. It is evident from this plot that a maximum of approximately 20 principal components is needed to represent the data without much loss of information, as the size of the eignevalues drop by an order of magnitude by the 20th principal component. The complete data set (including the training, validation and test sets) were shifted to the centroid of the original training set by subtracting the mean of the training set from each vector. It was then projected on to the new basis using the first 25, 20 and 15 principal components. Diagnostic systems based on multi layer perceptrons with the backpropagation algorithm were then developed using the data with reduced dimensionality maintaining the same proportion for training, validation and testing as discussed earlier. The best performance was achieved by the net with the 25 principal components (94.27%).

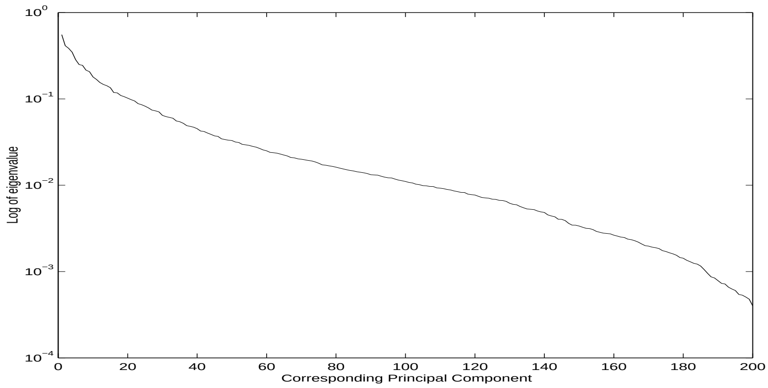

Figure 2: Log plot of eigenvalues corresponding to the principal components computed from the covariance matrix of the training data set.

4 Domain expertise

The vibration data was converted to digital form by an analog to digital (AtoD) card. Since the engine speed and load vary during routine operation, the data would be highly non-stationary in the time domain. To counter this problem, the AtoD card was driven using an external clocking signal from a shaft encoder which was synchronised with the rotary movement of the crankshaft. For more details on the data acquisition system, see Chandroth & Sharkey[2]. This approach made the data invariant to speed as each item of data was synchronised with the angular position of the crankshaft and consequently tied up with the engine events. The combustion process, when the air-fuel mixture is ignited, represents an area of violent activity in the engine cycle. Any malfunctioning component should cause changes, albeit minor, to the pattern of vibrations associated with the gas pressure oscillations during combustion. Sub-sampling the data in the 180^o to 200^o range by a factor of 4, computing the absolute values and normalising each data vector, a new set of data was created to train fully connected feedforward neural nets using backpropagation. After training and validating several nets, the best system was tested to an error tolerance of 0.1 on a test set of 3000 samples. The net produced a correct response to 90.23% of the test data.

5 Majority voting system

As discussed in the previous sections, several systems were developed using data from the same source but preprocessed in diverse ways. The best generalisation performance (97.4%) was achieved with the wavelets based preprocessing approach. It would be a waste of resources if only the best among these was selected and the rest discarded, as is the common practice. Further, it could be advantageous to fuse the decisions of several nets together in a majority voting system. This strategy would have the same effect as combining the decisions of several experts to form a robust and reliable classification. There have been several demonstrations of the improved performance that can result from combining a set of nets in an ensemble (for reviews see Sharkey[7] and Sharkey[8]). Clearly there would be no advantage to combining a set of identical nets. What is required is that the component nets in an ensemble should be both as accurate as possible and diverse, in the sense that they show different patterns of generalisation. Where the errors made by the component nets in an ensemble are independent, the errors made by one net would be compensated for by the correct outputs from the other nets in the ensemble, and, when combined by means of a majority vote could result in improved performance.

A variety of selection algorithms have been suggested for combining nets. The method adopted here was to exhaustively try out all combinations of nets from the 3 sets mentioned earlier and choose the best performing combination as the final system. In total there were 8, 3 and 5 nets from each of the methodologies (i.e Wavlelet, PCA and Domain expertise). There were 560 possible ways of combining these nets in combinations of 3. Out of the 560 combinations, 486 combinations performed better than each of their individual constituents when tested on the validation set. Figure **??** is a histogram of the results of testing. The best performance was 100% in 2 instances and the worst was 91.07% . When the best two ensembles with 100% generalisation performance were then tested on the test set, their respective results were 98.73 % and 95.83% .

An interesting observation is that the ensemble which attained 100% accuracy on the validation set and 98.73% on the test set, consisted of one net from each of the 3 methods. The wavelet net's performance was 90% the PCA net's was 93.73% and the domain expertise net's was 89.2% . Thus there was an average improvement in

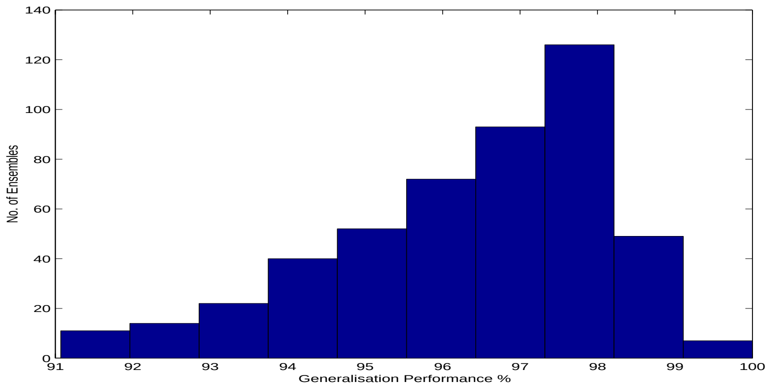

Figure 3: Histogram of generalisation performance of the 486 out of 560 ensembles of nets which performed better than each their individual constituents

performance of over 9% . These results add strength to the argument that it is desirable to combine nets exhibiting diversity rather than combine the best performing nets.

6 Conclusion

We have demonstrated that a single parameter (vibration), acquired using a non-intrusive accelerometer can be used to identify some commonly occurring faults in a diesel engine. By extracting features from the data using orthogonal wavelets, principal components and domain expertise, we showed how neural nets trained on these data performed to high levels of accuracy. In comparing the effectiveness of the 3 preprocessing methods adopted, we found that the best performance was achieved with the wavlets approach. Further, the technique of combining an odd number of nets into ensembles which make a decision on majority vote, has been shown to be a powerful method which not only increases the confidence in the decision of the system, but also avoids wastage by seeking out the most diverse set of nets from the entire pool of available nets.

References

[1] Bishop, C.M., *Neural Networks for Pattern Recognition,* Oxford University Press, Oxford, 1995.

[2] Chandroth, G.O., & Sharkey, A.J.C., Utilising the Rotational Motion of Machinery in a High Resolution Data Acquisition System, *Proc. of Computers and Ships- from Design and Build, through Automation and Management and on to Ship Support,* London, May 11- 12, 1999.

[3] Chui, Ch.K., *Wavelets: A Mathematical Tool for Signal Analysis,* SIAM, Philadelphia, 1992.

[4] Daubechies, I., *Ten Lectures on Wavelets,* SIAM, Philadelphia, 1992.

[5] Joliffe, I.T., *Principal Component Analysis,* New York, Springer-Verlag.

[6] Priede, T., Noise and Vibration Control of the Internal Combustion Reciprocating Engine, *Noise and Vibration Control Engineering: Principles and Applications,* ed. L.L.Beranek and I.L. Ver, John Wiley & Sons, Inc.

[7] Sharkey, A.J.C., (eds.), *Combining Artificial Neural Nets: Ensemble and Modular Multi-Net Systems,* Springer-Verlag, London, 1999.

[8] Sharkey, A.J.C., Combining Artificial Neural Nets, Connection Science 8(3/4), pp. 299-314, 1997.

[9] Staszewski, W.J., Ruotolo, R. & Storer, D.M., Fault Detection in Ball–Bearings Using Wavelet Variance, *Proc. of the 17th International Modal Analysis Conference (IMAC),*Kissimmee, Florida, 8-11 February, 1999.