

NASA
0008-SC-1002
CONF. PAPER
1N/82

APPENDIX F: Analysis of the Westland data set

Presented at the SPIE Aerosense Conference, Orlando, FL, April 16-20, 2001.

Analysis of the Westland Data Set

Fang Wen, Peter Willett

Electrical and Computer Engineering Department
U-2157, University of Connecticut
Storrs, CT 06269
(860) 486-2195, willett@engr.uconn.edu

Somnath Deb

Qualtech Systems, Inc.
100 Great Meadow Rd, Set. 501
Wethersfield, CT
(860) 257-8014, Deb@teamqsi.com

ABSTRACT

The “Westland” set of empirical accelerometer helicopter data with seeded and labeled faults is analyzed with the aim of condition monitoring. The autoregressive (AR) coefficients from a simple linear model encapsulate a great deal of information in a relatively few measurements; and it has also been found that augmentation of these by harmonic and other parameters can improve classification significantly. Several techniques have been explored, among these restricted Coulomb energy (RCE) networks, learning vector quantization (LVQ), Gaussian mixture classifiers and decision trees. A problem with these approaches, and in common with many classification paradigms, is that augmentation of the feature dimension can degrade classification ability. Thus, we also introduce the Bayesian data reduction algorithm (BDRA), which imposes a Dirichlet prior on training data and is thus able to quantify probability of error in an exact manner, such that features may be discarded or coarsened appropriately.

1. INTRODUCTION

Qualtech Systems has developed a suite of fault-isolation tools (TEAMS) that can, in real time and based on binary sensor data, isolate single and even multiple faults in complex systems. However, many sensors (for example, of vibration) are incapable of reliable decision-making on their own, and hence it has become necessary to develop a (real-time) signal processing “front-end” to the TEAMS inference engine, whose goal is to render single- and multiple-sensor level decisions as intelligently as possible. The signal processing system includes a wide menu of spectral and statistical manipulation primitives such as filters, harmonic analyzers, transient detectors, and multi-resolution decomposition.

The signal processing kit includes pattern classification software, including techniques based on restricted Coulomb energy (RCE), decision trees (DT), learning vector quantization (LVQ), fuzzy logic, Bayesian data reduction (BDRA), Gaussian mixtures (GM) and multi-layer perceptrons (MLP). At present the former three are implemented within the SP toolkit, and the fifth and sixth are implemented off-line in MATLAB using features provided by the toolkit.

Recognition of faults can hence be automated provided there is sufficient training data. This paper thus includes analysis of no-fault and seeded-fault vibration data from a CH-46 (“SeaKnight”) helicopter aft gearbox as collected from a test-stand. This data is made freely available through the generosity of the Penn State ARL.⁸

Results show promising fault detection accuracy, particularly when learning is based on auto-regressive (AR) coefficient features. The analysis presented in this paper is an outgrowth of that in.^{11,12} In the first of these, only a very abbreviated version of the Westland dataset was explored, and the RCE, LVQ, and DT schemes were discussed. In the second and this paper the full dataset is used, and the set of classifiers is augmented by the GM and BDRA approaches.

In section 2 we go into detail about the toolbox classification techniques: LVQ, DT, RCE, Gaussian mixture, and BDRA classifiers. In section 3 we apply the signal processing and classifiers to the Westland helicopter dataset. Similar to results reported elsewhere, we find near-perfect fault-recognition accuracy, in our case with relatively small feature sets involving autoregressive coefficients. We are encouraged by the success of the BDRA in its automatic digestion of the large number of features down to a relative few, and that these overlap significantly with the “accepted” ones.

2. THE CLASSIFIERS

We offer in the following a brief discussion of the SP toolbox's classification capabilities.

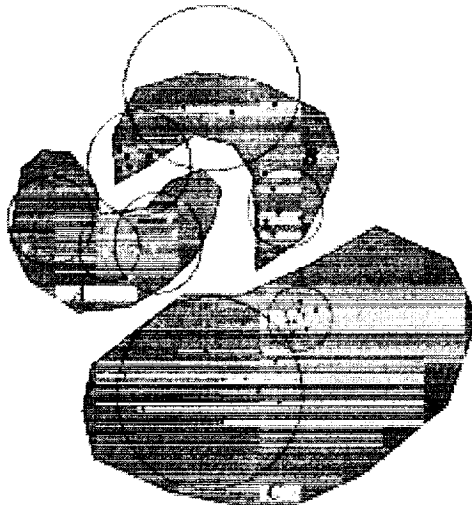


Figure 1. Illustration of classifier approximation via the RCE approach. The classes labeled A, B, and C, can generate observations within the indicated shaded regions. The actual training observations are given by dots or small squares. The approximating circles are shown also.

2.1. Restricted Coulomb Energy Classification

At fundament, the RCE classifier^{4,9} relies on the approximation of a decision region via a union of hypersphere “cells”, as illustrated in two dimensions in figure 1. Cells may overlap if they do not belong to the same class, and this may produce ambiguous outputs. Note that partition of the observation space into decision regions is not exhaustive in the RCE approach.

Training of an RCE classifier is of course iterative, with the training data sets cycled repeatedly as training “epochs”. Training is as follows:

1. Randomly shuffle training data.
2. Consider a training data point, and find those hyperspheres which contain it.
 - (a) If there are none, then initialize a new cell centered at that data point and having a (pre-specified) maximum radius.
 - (b) If a containing hypersphere is associated with the correct class, then do nothing.
 - (c) If a containing hypersphere is of an incorrect class, then decrease its radius to correct this.
3. Repeat the above for all data in training set.
4. If the training epoch has resulted in changes neither in hypersphere membership nor in modification of hypersphere number or radius, stop. Otherwise return to 1.

We are not aware of a proof of convergence of RCE training, but we have observed no lack of convergence.

After the network has become fixed classification is accomplished by interrogation of membership of the various cells: each cell is assigned a class, and the output corresponds to that class. For the cases that data is either a member of no cell, or of several which are of different classes, the RCE classifier gives an indeterminate output: such cases may be decided randomly or by heuristic.

The RCE classifier appears to be a good choice when the classes are separable (i.e. an ideal classifier would operate without error) and when there is sufficient training data that separation is possible. This is similar to simpler setups such as linear and quadratic discriminant analysis¹⁰; but whereas those techniques must have decision boundaries either hyperplanar or hyperellipsoidal shells, the RCE decision regions can be quite weirdly-shaped and non-convex.

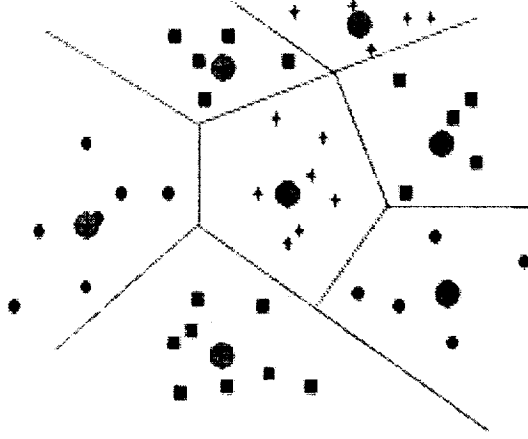


Figure 2. Illustration of classifier approximation via LVQ approach. Training observations are denoted as stars, boxes, or circles, depending on class. Subclusters are formed for each class, and the centroids for each are represented as \oplus .

2.2. Learning Vector Quantization Classification

The LVQ classifier⁵ is a variation on the traditional cluster-classifier based on K-means training.¹⁰ In essence, each class is assigned sub-clusters defined by their centroids (see figure 2), and data are classified based on the membership of the centroid to which they are nearest.

Training of an LVQ classifier must proceed from an initial “guess” at cluster centroids; this may be from K-means. Thereupon consider training datum x_i , which is of class $C(x_i)$ and is closest to centroid μ_j : if the membership of μ_j is also $C(x_i)$, then we update in the direction of the new data

$$\mu_j^{\text{new}} = \mu_j^{\text{old}} + \eta(x_i - \mu_j^{\text{old}}) \quad (1)$$

and otherwise in the opposite direction

$$\mu_j^{\text{new}} = \mu_j^{\text{old}} - \eta(x_i - \mu_j^{\text{old}}) \quad (2)$$

Generally η is decreased from epoch to epoch. Eventual convergence is assured here provided η is sufficiently small, but in practice training is ceased when changes become insignificant.

In our implementation clusters are never created, but may be merged. After cessation of training (as above), successive pairs of common-class clusters are testing for the appeal of their combination: for the proposed super-cluster a new mean μ^* and “radius” R^* are calculated, the former being the usual centroid and the latter the greatest Euclidean distance from the centroid to a member point. If this radius is less than a distance β to the nearest centroid not in the current class, then the merge is accepted, and otherwise it is not. Typically we have $.33 \leq \beta \leq 1$.

An LVQ classifier may be considered a development on the earlier K-means based cluster classifiers in that non-separable classes cause no intrinsic, and in that there is an intelligent means of “pruning” clusters.

2.3. Decision Tree Classifier

At core the DT classifier¹⁰ produces its output by asking a series of questions which must have binary answers. These queries, for instance “Is feature 2 greater than 7.53?” may be based on previous answers, and each must interrogate one feature alone. The “path” taken may be thought of as traversal of a logical tree; but the form of the resultant decision regions must be as hyper-rectangles, as illustrated in figure 3.

In principle it is possible and easy to separate the training data precisely via a sufficiently-rich question set. In practice there are too many “questions” (parameters), and the DT classifier is found not to have a particularly good generalization ability. There are a number of means to limit the number of questions, and these generally amount to the choice of a cost to be placed on a question’s posing. In our implementation we use an information-theoretic cost function, although admittedly its basis is empirical rather than true prior statistics.

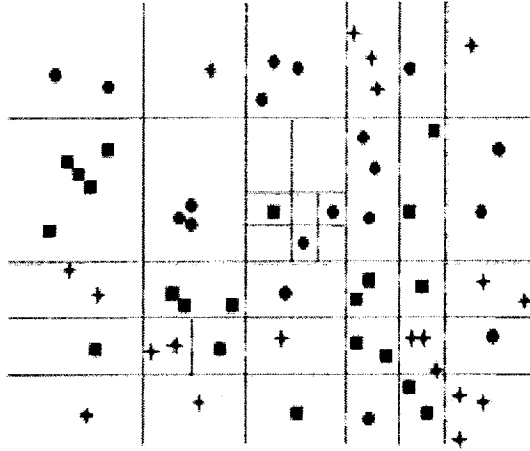


Figure 3. Illustration of classifier approximation via DT approach. Training observations are denoted as stars, boxes, or circles, depending on class. Subclusters are based on answering yes/no questions, and hence must be hyper-rectangular – in two dimensions, they must be rectangles.

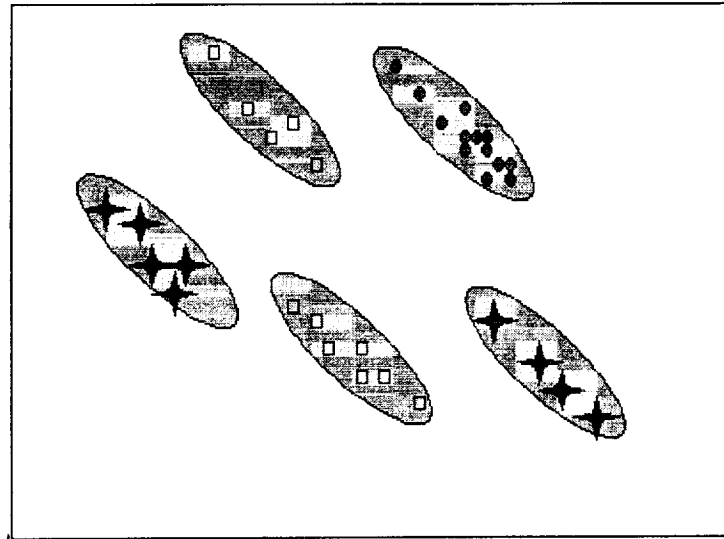


Figure 4. Illustration of classifier approximation via Gaussian mixture approach. Training observations are denoted as stars, boxes, or circles, depending on class. Ellipses denote regions of support of Gaussian mixture elements.

2.4. Gaussian Mixture Classifier

This classification technique has a greater statistical grounding than the previous, in that a probability density function (pdf) is sought for each class. The specific pdf used is a mixture of multivariate Gaussians:

$$f(\mathbf{x}) = \sum_{i=1}^M \pi_i \frac{1}{|2\pi\mathbf{R}|} e^{-\frac{1}{2}[\mathbf{x}-\mu_i]^T \mathbf{R}^{-1}[\mathbf{x}-\mu_i]} \quad (3)$$

and the idea is illustrated in figure 4. There are M elements to the mixture, and each has a different mean μ_i and prior probability π_i . Decisions are made according to the maximum posterior probability of each class (in fact, classes are assumed to be equally-likely a-priori). Note that if $M = 1$ this is identical to the quadratic discriminant classifier.

Training is via the expectation/maximization (EM) algorithm.¹⁰ The correlation matrix \mathbf{R} is common to all elements of the mixture within a class of fault – this is known as a “homoscedastic” mixture – and the ideas behind this are that the number of elements to be estimated can be reduced and that there is little concern about unboundedness of the likelihood function. A variant of the above restricts \mathbf{R} to be diagonal; this reduces the number of parameters to estimate considerably, but in this particular case (see, for example, figure 6) the ability to “tilt” the pdf level curves arising from the use of a full \mathbf{R} is valuable.

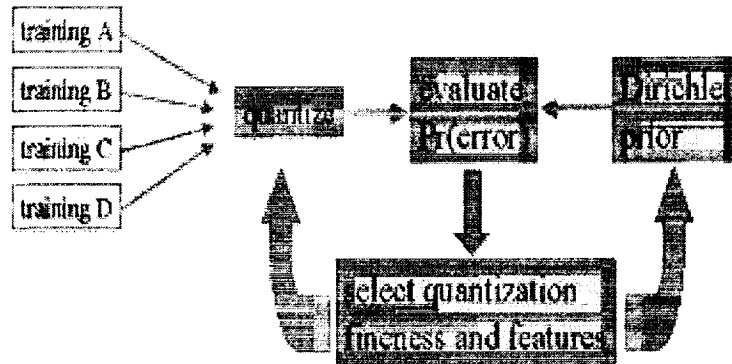


Figure 5. Illustration of approach to classification via BDRA.

2.5. BDRA Classification

The Bayesian data reduction approach is perhaps the most statistically defensible of the classifiers used. It begins with a quantized version of the data, and assumes a Dirichlet prior (of complete ignorance) on this a priori, for each fault class. From that prior distribution classification is relatively simple; the key is that the prior enables an explicit (and correct) probability of error to be calculated, and thence features may be pruned in an optimal way – this is illustrated in figure 5. The BDRA is discussed in detail in,⁶ among other places. Generally the BDRA works very well when there are too many features for the training data to support, and/or when the classes are not easily separable.

The BDRA requires that the data be pre-quantized. To some extent this is not a concern, since the quantization may be as fine as desired – the BDRA coarsens the quantization as part of its feature/level selection. For practical reasons, the quantization cannot be too fine, and hence it is not expected that this dataset will be kind to the BDRA. In fact, the BDRA results are reasonable, but what is interesting is its ability to select features and its prediction of its own performance.

3. RESULTS

3.1. The Data

In the early 1990's the US Navy contracted with Westland, a British helicopter manufacturer, to develop and study vibration signatures for the CH-46 (SeaKnight) aft gearbox. Essentially this is “test-stand” (not in-flight) data; this is a disadvantage from the perspective of result reliability, but offers a distinct advantage in that the vibration signatures are labeled. The data is as follows:

- There are 68 files each containing data traces of 100,000 samples.
- For each case there is data available from eight accelerometers.
- There are a total of nine fault conditions, ranging in severity from mild to severe. Faults were “seeded” (by electronic discharge milling) in the sense that parts with known defects were installed and de-installed.
- There is data from no-fault (normal) operating conditions.
- Data was observed at nine different torque levels (since this is a rotorcraft, angular velocities are relatively constant), ranging from 27% to 100%.

For details on the faults, etc., please see.^{1,8} Note that if all fault levels and torques were represented there would be 90 files; in fact, a number of conditions are unrepresented in the data. As regards training versus testing, the entire dataset is split randomly into two parts, which are used separately.

The data has been analyzed previously (e.g.^{17,13}) using a variety of classification techniques such as multi-layer perceptrons and fuzzy reasoning. Indeed, this is apparently an “easy” (or separable) dataset for classification, as the reported accuracies approach 100%. Thus, our goal here is not really to beat previous (unbeatable!) approaches, but to attempt to match them using the SP toolbox classifiers. Further, it appears that past approaches have often used a rather dense feature set (several hundred features, such as FFT outputs), and we attempt here to use a much sparser arsenal.

3.2. The Features

3.2.1. AR Coefficients

It is possible to use periodogram outputs explicitly as features for classification; however, in general this implies a great many features, and the usual “curse of dimensionality” may ensue. Since it is clear that spectral features do indeed yield much relevant information, we propose to use a concise way of representing the spectrum: the autoregressive (AR) parameters.² These are estimated on blocks of various sizes, from $N = 256$ to $N = 16384$.

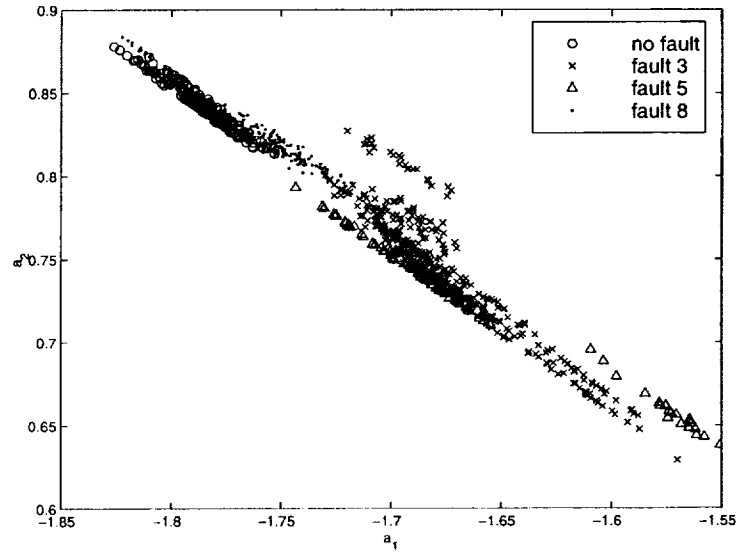


Figure 6. Scatter plot of AR coefficients a_1 versus a_2 (out of $p = 4$ AR coefficients), for accelerometer 3 and combined over all torque levels, estimated on blocks of length 4096, for faults 3, 5, 8 and no-fault conditions.

Examples of AR coefficients are given in figures 6 and 7. It is clear that there is a reasonable amount of structure to these, but also that certain conditions cannot be separated reliably using only such data. In fact, there are 8 accelerometers from which to choose, and a further two AR coefficients.

3.2.2. FFT Features

AR coefficients are able to digest much global spectral information into a small dimension. There is some indication that faults may manifest in specific frequency behavior, and hence we additionally investigate the use of relative harmonic power (RHP). The i^{th} RHP is the ratio of the i^{th} -highest spectral peak (measured via FFT) to the average power. In the sequel we use 4 RHP’s. The idea is illustrated in figure 8, and examples are given in figures 9 and 10, for the same conditions as figures 6 and 7. It is clear that these features are less a direct indication of fault class.

3.3. Results for RCE, LVQ and DT

We first examine the results for the case that accelerometers are used individually. The features used are AR coefficients of order $p = 4$, each estimated on a block of length $N = 4096$. Results are reported in table 1. None of these performances is acceptable, although accelerometers 3, 4, and 7 appear to be the most promising. Motivated

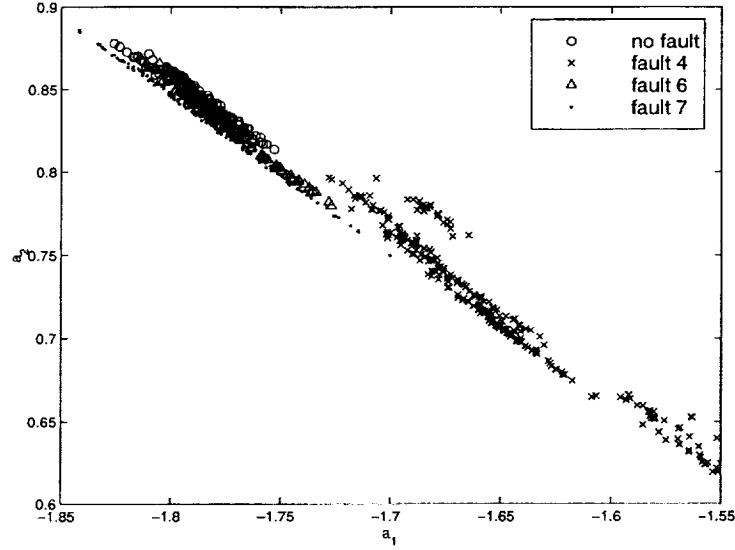


Figure 7. Scatter plot of AR coefficients a_1 versus a_2 (out of $p = 4$ AR coefficients), for accelerometer 3 and combined over all torque levels, estimated on blocks of length 4096, for faults 4, 6, 7 and no-fault conditions.

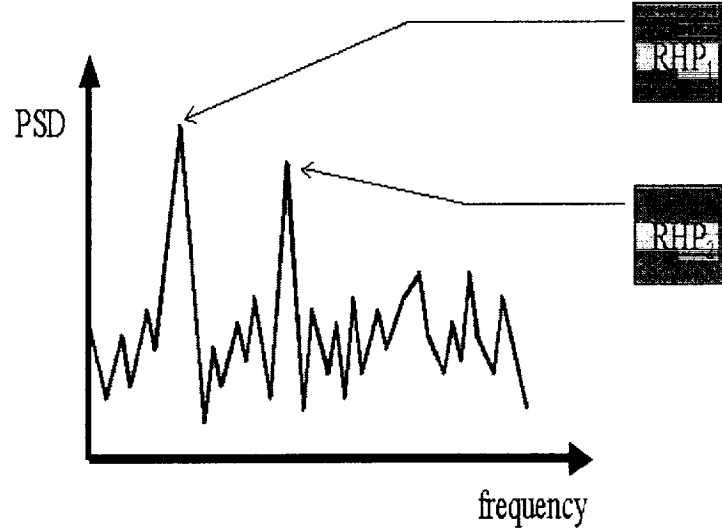


Figure 8. Illustration of extraction of relative harmonic power coefficients.

by this, we attempt to classify by combining accelerometers. Example results are shown in table 2. We find that the combination of accelerometers 3 and 7 is the most propitious. There is apparently little benefit from using all accelerometers.

In table 3 we explore the choice of AR order. The results indicate that $p = 4$ is a good compromise between sensitivity and dimensionality. With this choice we consider adding the RHP features. In table 4 we do, and additionally compare the results for different block lengths. The results become quite outstanding in the cases $N = 4096$ and $N = 16384$, particularly for the RCE classifier; the LVQ classifier is somewhat less satisfying, and the DT scheme has been overcome.

Finally, we note that we have chosen to ignore the torque level in our classification feature set. That is, we have trained using combined data from all torque levels, and results to this point are given in terms of combined probability of correctness. It could be argued that this is dangerous, in that poor performance may lurk at some

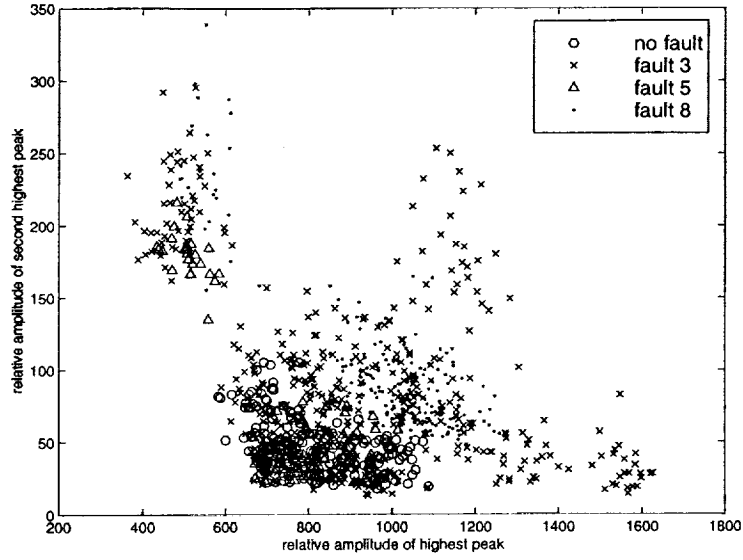


Figure 9. Scatter plot of RHP values, corresponding to figure 6.

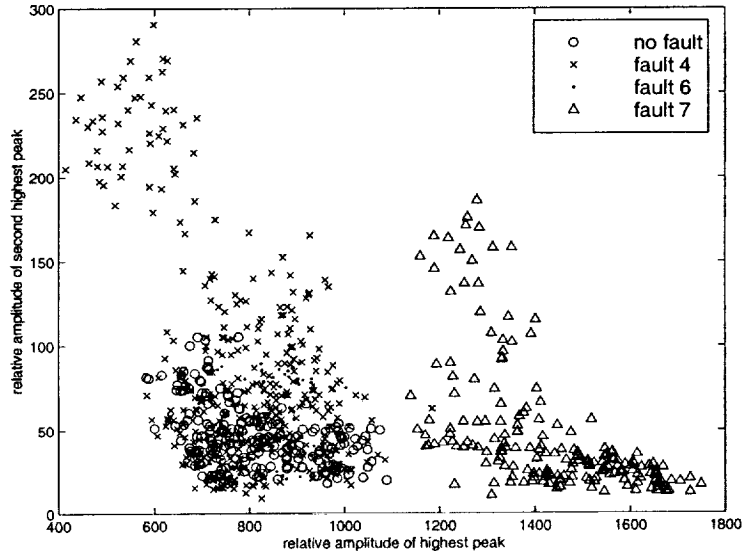


Figure 10. Scatter plot of RHP values, corresponding to figure 7.

torque level; in fact, as seen in table 5, this is not the case.

3.4. Results for GM

We show example results for the two homoscedastic GM classifier variants in figure 11. Apparently the GM classifier is not as good as the RCE scheme in this situation; GM classifiers are often more useful when the data is less separable and when confidence information is desired, so it is perhaps interesting that the performance is as good as it is. Of particular note is the $M = 1$ GM classifier – this is essentially a quadratic discriminant, and its probability of error is very low.

As regard the second GM classifier variant – that with a diagonal covariance matrix – it is interesting to observe from figure 11 that the performance improves markedly as the number of mixture elements M is increased. There is some explanation of this in figure 12, in which the “coverage” of one class’s data by the mixture elements is illustrated. It is clear that the more elements, the more complete the coverage.

acc	RCE	LVQ	DT
1	54	29	52
2	49	76	72
3	78	27	73
4	78	38	72
5	60	33	53
6	57	32	48
7	84	34	81
8	30	14	31

Table 1. Percentage of correct classification for three classifiers, versus accelerometer number – features are 4th-order AR coefficients from individual accelerometers.

acc	RCE	LVQ	DT
2,7	96.4	89.4	94.7
3,7	99.1	96.7	95.5
4,7	98.3	89.6	94.9
all	98.5	96.6	91

Table 2. Percentage of correct classification for three classifiers, with combined accelerometer AR coefficients ($p = 4$) from individual accelerometers as features. (In the last row $p = 2$.)

3.5. Results for BDRA

In table 6 we show the results for the BDRA in terms of correct detection of a fault condition – no attempt is made here to isolate the fault, but testing is simply binary. Despite the fact that the BDRA is not particularly well-suited to the problem, the results are quite good. It is particularly notable that the algorithm is able to predict its own performance with reasonable fidelity.

As indicated earlier, a strength of the BDRA is that it is able to determine for itself a feature set. In fact, it is originally “given” a the entire set of features, quantized to whatever fineness is desired – in table 6 this is 5 or 10 levels per feature, thresholded for equal probability, meaning in the case of 10 levels and $p = 6$, there are initially $8 \times (6 + 4) \times 10 = 800$ possible observations. In table 7 the final quantization from the BDRA is shown, and the dominance of accelerometers 3 and 7 is clear. Table 7 deals only with AR coefficients: if RHP features are also presented to the BDRA, it turns out that these are often chosen.

The BDRA is capable of multi-class operation, although in the paper¹² that shares much in common with this one, this feature was not implemented there. In table 8 we see results for this operation (with mixed-torque data), and it is clear that the BDRA is able to ascertain the error classes with surprising accuracy. However, note from the earlier figures 6 and 7 that the AR coefficients are in fact strongly correlated with each other, and tend to cluster in highly eccentrically elliptical regions – the BDRA is not particularly suitable for this sort of case. However, by a simple “whitening” pre-processing operation (actually, estimation of a common correlation matrix \mathbf{R} for all clusters, similar to the Gaussian mixture classifier with $M = 1$ mode, and multiplication of the entire data set by $\mathbf{R}^{-1/2}$), the performance is clearly much improved. A confusion matrix is shown in table 9, and performance clearly is reasonable: the exceptions here relate to class 3 (in fact, “Input Pinion Bearing Corrosion”) which appears to be similar to many other faults; and to class 2 (“Planetary Bearing Corrosion”) which is so under-represented in the data that it cannot attract a decision.

p	RCE	LVQ	DT
2	83.2	88.7	87.7
3	94.0	85.4	97.4
4	99.1	96.7	95.5
6	99.0	92.8	95.0
8	98.9	91.1	95.9
12	97.6	96.1	93.8

Table 3. Percentage of correct classification for three classifiers, with combined accelerometers 3 and 7, for various AR orders p , estimated on data blocks of length $N = 1024$.

N	RCE	LVQ	DT
1024	96.8	95.1	93.1
4096	99.6	97.9	93.9
16384	99.3	96.7	91.9

Table 4. Percentage of correct classification for three classifiers, with combined accelerometers 3 and 7, for various AR orders $p = 4$ estimated on data blocks of length N . The feature set is augmented by the RHP spectral peak clues.

torque	RCE	LVQ	DT
27%	97.6	94.1	92.9
40%	100	100	98.8
45%	100	100	100
50%	100	100	97.2
60%	100	98.6	88.9
70%	99.1	82.4	83.3
75%	97.9	99.0	84.8
80%	99.1	98.2	91.7
100%	100	98.3	89.2

Table 5. Percentage of correct classification for three classifiers, with combined accelerometers 3 and 7, for various AR orders $p = 4$ estimated on data blocks of length $N = 4096$. The feature set is augmented by the RHP spectral peak clues. Training data is combined over all torque levels, and testing is done individually at each torque level.

4. SUMMARY

Here we have reported on a signal processing toolbox specially matched to Qualtech Systems TEAMS diagnostic inference engine, and in particular on its classification capability as applied to the “Westland” data set. We have found that essentially perfect diagnostic performance is achievable via the use of AR coefficient features augmented by harmonic peak information. The best classification performance appears to come from the RCE learning/classification scheme. The approach works well across all torque levels, so there is no need to supply engine load information to the classifier. We have also found that the Bayesian data reduction (BDRA) approach, despite not being well-matched to the problem, works surprisingly well, and indeed that its ability to select features (perhaps for other classifiers?) is particularly promising.

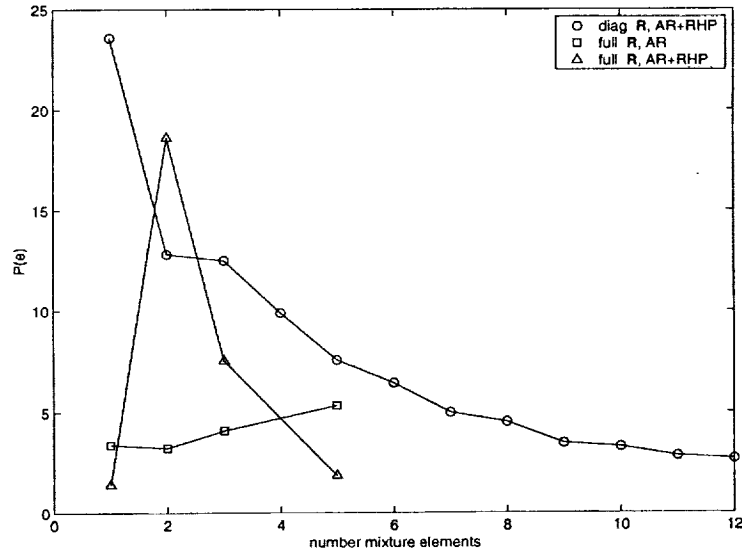


Figure 11. Probability of error performance for Gaussian mixture classifiers. Here $p = 4$ and $N = 1024$.

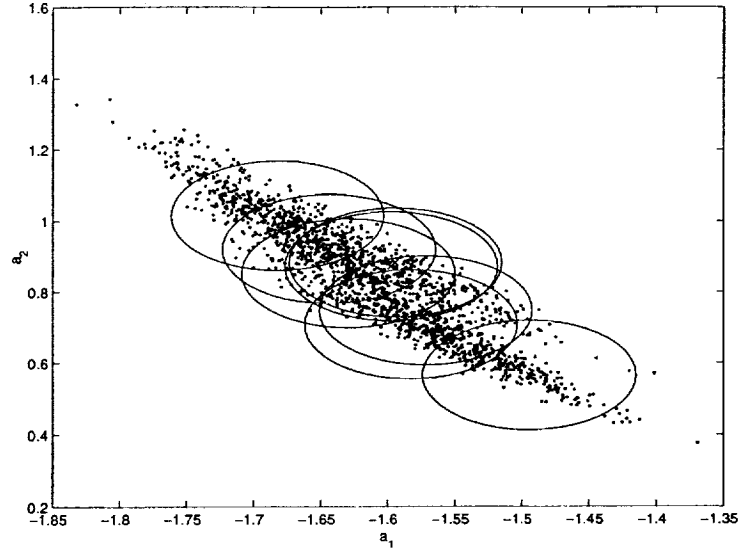


Figure 12. Dots show scatter plot of AR coefficients a_1 versus a_2 (out of $p = 4$ AR coefficients), for accelerometer 3 and combined over all torque levels. Ellipses are probability contours for elements of diagonal- \mathbf{R} Gaussian mixture fit with 8 elements.

N	p	C_5^t	C_5^a	C_{10}^t	C_{10}^a
1024	2	93.9	95.7	95.3	96.5
1024	4	94.9	95.4	94.3	94.9
1024	6	94.5	96.3	94.0	94.1
4096	2	93.9	98.9	96.5	98.8
4096	4	94.4	96.3	95.4	98.3
4096	6	92.2	91.3	93.8	98.0

Table 6. Percentage of correct fault detection for BDRA, using $AR(p)$ coefficients and RHP clues. Subscript of C denotes number of initial quantization levels per feature; superscript a means actual, and t is theoretical.

5. ACKNOWLEDGMENTS

This research was supported by ARO STTR contract DAAG55-98-C-0057 and NASA contract 0008-SC-1002, each with prime contractor Qualtech Systems, Inc., and a subcontract to the University of Connecticut. The authors also wish to acknowledge the help of Vivek Rajan for his work on the classification package, and the support of NASA. The authors are grateful to Penn State ARL for making this data freely available on the world wide web.

REFERENCES

1. B. Cameron, "Final Report on CH-46 Aft Transmission Seeded Fault Testing", Westland Helicopters Research Paper RP907, Westland Helicopters, 1993.

N	p	f_1	f_2	f_3	f_4
1024	2	a_3^1	$2 \times a_7^2$	—	—
1024	4	a_3^4	a_4^4	a_7^3	$2 \times a_7^4$
1024	6	a_2^6	a_4^5	a_7^1	$2 \times a_7^3$
4096	2	a_3^4	$2 \times a_7^2$	—	—
4096	4	a_4^4	a_7^4	—	—
4096	6	a_2^5	a_3^4	$2 \times a_7^1$	—

Table 7. Features used by BDRA; only AR features present, initially quantized to 10 levels. The notation a_i^j means that the j^{th} AR coefficient of accelerometer i is active, and $2 \times a_i^j$ means that three levels (two thresholds) are kept.

N	p	C_5^t	C_5^a	C_{10}^t	C_{10}^a	W_5^t	W_5^a	W_{10}^t	W_{10}^a
1024	2	76.3	81.3	76.7	82.4	80.4	88.4	85.0	89.6
1024	4	70.4	76.0	73.1	83.9	80.7	84.9	82.0	87.7
1024	6	73.4	78.8	72.5	78.6	86.2	91.0	87.8	92.3
4096	2	81.0	88.8	78.6	86.6	78.2	86.0	79.7	90.0
4096	4	70.1	82.2	72.4	86.5	80.1	89.1	78.9	83.8
4096	6	70.3	80.4	70.7	76.8	81.0	86.9	83.7	90.7

Table 8. Percentage of correct fault isolation (i.e. multi-class) for BDRA. Columns headed by C use AR(p) coefficients and RHP clues. Subscript of C denotes number of initial quantization levels per feature; superscript a means actual, and t is theoretical. Columns headed by W use whitened AR(p) clues only.

	2	3	4	5	6	7	8	9
2	0	0	0	0	0	0	0	0
3	12	204	13	1	7	1	22	0
4	0	12	155	0	0	0	0	0
5	0	0	9	46	0	0	0	0
6	0	0	0	1	32	0	0	0
7	0	0	2	0	0	95	0	1
8	0	0	0	0	1	0	86	0
9	0	0	0	0	8	0	0	107

Table 9. Confusion matrix for multi-class BDRA with FFT length 4096 and $p = 2$ AR features. The $(i, j)^{\text{th}}$ entry denotes the number of decisions for class i when the true class is j . Class 9 is “no fault”, and there is no data available from class 1.

2. S. Haykin, **Adaptive Filter Theory**, 3rd ed., Prentice Hall, 1996.
3. M. Horsey, “WHL Vibration Analysis Results from CH46 Aft Gearbox Testing”, **Westland Helicopters Mechanical Research Report MRR 20168**, 1993.
4. M. Hudak, “RCE Classifiers: Theory and Practice”, **Cybernetics and Systems: An International Journal**, vol.23, pp483-515, 1992.
5. T. Kohonen, J. Kangas, J. Laaksonen, and K. Torkkola, “LVQ-PAK: A Program Package for the Correct Application of Learning Vector Quantization Algorithms”, **Proceedings of the IJCNN**, pp725-730, 1992.
6. R. Lynch and P. Willett, “Performance Considerations for a Combined Information Classification Test Using Dirichlet Priors”, **IEEE Transactions on Signal Processing**, pp1711-1714, June 1999.
7. P. Monson, M. Dwonczyk, and E. Manolakos, “Analog Neural Networks Based Helicopter Gearbox Health Monitoring System”, **Journal of the Acoustical Society of America**, vol.96(6), pp3235-3249, 1995.
8. Pennsylvania State University Applied Research Laboratory, <http://wisdom.arl.psu.edu/Westland/>.
9. D. Reilly, “The RCE Neural Network”, **CRC Press Industrial Handbook**, pp.1025-1037, 1997.
10. B. Ripley, **Pattern Recognition and Neural Networks**, Cambridge University Press, 1997.
11. F. Wen, P. Willett, and S. Deb, “Signal Processing and Fault Detection with Application to CH46 Helicopter Data”, **Proceedings of the 2000 Aerospace Conference**, Big Sky MT, March 2000.
12. F. Wen and P. Willett, “Condition Monitoring for Helicopter Data”, **Proceedings of the 2000 Conference on Systems, Man, and Cybernetics**, pp. 224-229, Nashville TN, October 2000.
13. G. Yen, “Health Monitoring of Vibration Signatures in Rotorcraft Wings”, **Neural Processing Letters**, vol.4(3), pp127-137, 1996.