ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 313-318

Feature Selection for ANNs using Genetic Algorithms in Condition Monitoring

L. B. Jack, A. K. Nandi

Department of Electrical Engineering and Electronics, The University of Liverpool, Liverpool, L69 3GJ, UK

1. Abstract

Artificial Neural Networks (ANNs) can be used successfully to detect faults in rotating machinery, using statistical estimates of the vibration signal as input features. One of the main problems facing the use of ANNs is the selection of the best inputs to the ANN, allowing the creation of compact, highly accurate networks that require comparatively little preprocessing. This paper examines the use of a Genetic Algorithm (GA) to select the most significant input features from a large set of possible features in machine condition monitoring contexts. Using a large set of 156 different features, the GA is able to select a set of 6 features that give 100% recognition accuracy.

2. Introduction

As modern day production plants are expected to run continuously, industry has created a demand for techniques that are capable of recognising the development of a fault condition within a machine system. Machine Condition Monitoring was developed to meet this need. Research has shown that ANNs are promising solution to several different problem areas[1, 2, 3]; however many of the input features used require a significant computational effort to calculate. A feature selection process using GAs is used in order to isolate features that are provide the most significant information for the neural network, whilst cutting down the number of inputs required for the network.

The work presented in this paper is based around experimental results performed on vibration data taken from a small test rig (see figure 1) which was fitted with a number of interchangeable faulty roller bearings. This is used to simulate the type of problems that can commonly occur in rotating machinery. Rolling element, or ball bearings, are one of the most common components in modern rotating machinery; being able to detect accurately the existence of a fault in a machine can be of prime importance in certain areas of industry.

Six different conditions are used within the experiments conducted for this paper. Two normal conditions exist: one bearing in a brand new condition,

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 313-318

while another is a bearing in slightly worn condition. There are four fault conditions:- inner race fault, outer race fault, rolling element fault, and a cage fault.

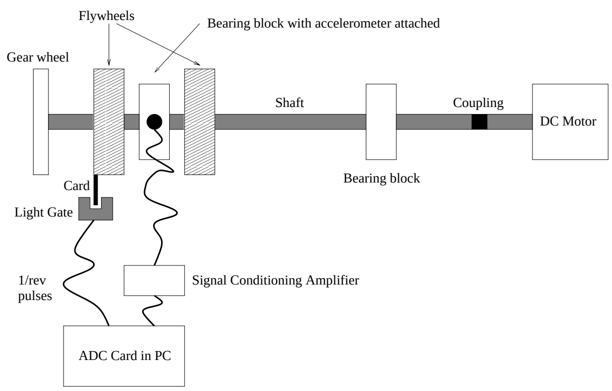

Figure 1: The machine test rig used in experiments

3. Signal Acquisition

The signal from the accelerometers was passed through a low pass filter with a cutoff frequency of 18.3 kHz, and was then sampled at a rate of 48 kHz, giving a slight oversampling of the data. This operation was repeated ten times for 16 different speeds. With a total of six different conditions, this gives a total data set of 960 cases, with 160 cases for each condition.

3.1. Vibration Datasets

A number of different statistical features were taken based on the moments and cumulants of the vibration data. Higher order spectra have been found to be useful in the identification of different problems in condition monitoring [4]. A good introduction is given in [5]. Using the measured signals, three datasets were prepared. The first dataset contained moments and cumulants of up to order 4 - a matrix of 960 entries.

A spectrally based feature set was created. For each of the two channels sampled, a 32-point FFT of the raw data was carried out, and 33 values were obtained for each channel. These were then stored as a column vector of 66 values, which was used as the input data set for the given data sample. The full input data set formed a 66×960 data set.

Combining these two datasets gave a larger third set with a total of 156 features (156×960). This dataset, and the other two smaller feature sets were all normalised before training.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 313-318

4. Neural Networks

The MLP used in this work consists of one hidden layer and an output layer, the hidden layer having a logistic activation function, whilst the output layer uses a linear activation function. The size of the hidden layer is determined by the genetic algorithm itself during training. This allows training to proceed at a faster rate than an exhaustive training process that checks different sizes of the first layer. The size of the second layer is determined by the number of outputs required. This is set at six neurons for the particular application. Training of the network is carried out using a standard back-propagation algorithm, and the network is trained for 350 epochs, using 40% of the data set as training data, and the remaining 60% as the test and validation set.

5. Genetic Algorithms

GAs have been gaining popularity in a variety of applications which require global optimisation of a solution. A good general introduction to genetic algorithms is given in [6]. The prime component of a genetic algorithm is the genome. The genome is an encoded set of instructions which the genetic algorithm will use to construct a new model or function (in this case the inputs to a neural network). The best type of encoding is very much problem dependent, and may require some form of combination of two or more encoding types (binary, real numbers, etc) in order to get the optimum results.

The GA is allowed to select subsets of various sizes in order to determine the optimum combination and number of inputs to the network. The emphasis in using the genetic algorithm for feature selection is to reduce the computational load on the training system while still allowing near optimal results to be found relatively quickly.

5.1. Feature Selection & Encoding

Feature selection of the GA is controlled through the values contained within the genome generated by the GA. On being passed a genome with $(N+1)$ values to be tested, the first N values are used to determine which rows are selected as a subset from the input feature set matrix. Rows corresponding to the numbers contained within the genome are copied into a new matrix containing N rows. The last value of the genome determines the number of neurons present in layer 1 of the network.

For this particular application, a simple real number based genome string was used. For a training run requiring N different inputs to be selected as a subset of Q possible inputs, the genome string would consist of $(N+1)$ real numbers. The maximum number of neurons permissible in the first layer is defined as S. Each of the first N numbers (x) in the genome is constrained as $0 \leq x \leq (Q-1)$, whilst the last number, (x), is constrained to lie within the bounds $1 \leq x \leq S$. This means that any mutation that occurs will be

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 313-318

bounded within the limits set at the definition of the genome. The classification performance of the trained network using the whole dataset was returned to the GA as the value of the fitness function.

The GA uses a population size of 10 individuals, starting with randomly generated genomes. The probability of mutation was set to 0.2, whilst the probability of crossover was set to 0.75. An elitist population model is used, meaning that the best individual in the previous population is kept in the next population, and preventing the performance of the GA worsening as the number of generations increase.

6. Training and Simulation

Training was carried out using three data sets; One feature set comprised all the statistically based features (90 features). The set of 66 spectral features was used as an individual case, and this dataset was combined with all the statistical feature sets to form an input feature set of 156 inputs. Each feature set contained a total of 960 cases. Using the genetic algorithm running for a total of 40 generations, each containing 10 members (meaning the training of 400 neural networks), eight separate cases were tested using various numbers of inputs, varying from five to twelve.

As a comparison, a neural network was trained using each feature set. These were trained for a total of 350 epochs, and allowed to choose the best size of intermediate layer between 2 and 15 neurons.

7. Results

7.1. Results: ANN

Table 1 shows a summary of results for all three feature sets used. The "No. Neurons" quoted in the second column is the number of neurons used in the hidden layer of the best network in each training run. "Classification success" represents the percentage success rate of the ANN using the complete dataset, which includes both training and test data. The "false alarm" rate details the number of "normal" conditions that were misclassified as alarm conditions, expressed as a percentage of the total dataset. The "fault not recognised" category details the number of fault conditions that were classified as normal, again expressed as a percentage of the total dataset.

As can be seen, all the feature sets have a performance greater than 88%, with two cases in excess of 91%; the spectral feature set gives the best performance, at 97% for the overall data set. While this is a comparatively small feature set containing 66 features, the spectral content of the data is ideally suited to recognising several of the periodic type faults that are generated by the different conditions. The aggregate of the false alarm rate and fault not recognised rates is also the lowest of all the different feature sets.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 313-318

Data Set	No. Inputs	No. Neurons	Classification Success (%)	False Alarm Rate (%)	Fault Not Recognised (%)
All Statistics	90	7	88.1	0.2	7.4
Spectral Data	66	8	97.0	0.2	2.2
All Data	156	6	91.1	0.1	10

Table 1: Performance of straight ANNs using different feature sets

	Straight ANN			GA with Best ANN			GA with ANN	
Data Set	No. I/Ps	No. Hid.	Perf. %	No. I/Ps	No. Hid.	Perf. %	Mean Perf. %	Perf. Range
All Statistics	90	7	88.1	7	14	97.7	97.1	96.0 - 97.7
Spectral Data	66	8	97.0	6	11	99.8	99.2	97.3 - 99.8
All Data	156	6	91.1	6	9	100	98.3	95.0 - 100

Table 2: Comparison between standalone ANN and GA with ANN after 40 generations, for all three data sets

The feature set containing all the data has a larger number of features than any of the other sets, and the drop in performance may be due to the limited amount of time that the training algorithm is allowed to arrive at a result. It may be that due to the large number of features the training algorithm is unable to arrive at a better solution before the training cycle is stopped.

7.2. Genetic Algorithm with ANN after 40 Generations

Table 2 shows the performance of the different feature sets after running under the GA for 40 generations. All of the datasets have their best performance in excess of 97.5%. The feature set using all the available training data has managed to achieve an accuracy of 100%, indicating accurate classification. This is achieved using only six inputs out of the possible 156. Using 9 neurons in the hidden layer, a relatively small network has been created that fulfils the criteria set earlier on. A network of this size would be ideal for a realtime implementation on a small chip or micro-controller.

ESANN'1999 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 21-23 April 1999, D-Facto public., ISBN 2-600049-9-X, pp. 313-318

8. Conclusions

The use of the Genetic Algorithm allows feature selection to be carried out in an automatic manner, meaning that input combinations can be selected without the need for human intervention. This technique offers great potential for use in a condition monitoring environment, where there are often hundreds and even thousands of different measurements available to a monitoring system, and selection of the most relevant features is often difficult. It has been shown that the Genetic algorithm is capable of selecting a subset of 6 inputs from a set of 156 features that allow the ANN to perform with 100% accuracy. The performance of networks trained using the feature selection was consistently higher than those trained without feature selection.

9. Acknowledgements

Thanks must be expressed to Weir Pumps for the loan of the machine set used in the experiments. Financial support was provided by Weir Pumps, Solatron Instruments, and the University of Liverpool.

10. References

References

[1] McCormick A. C. and Nandi A. K., Classification of Rotating Machine Condition using Artificial Neural Networks, Proceedings of the Institute of Mechanical Engineers, Part C, Vol 11, No. 6, pp 439-450, 1997.

[2] Sorsa, T. Koivo H. and Koivisto H., Neural Networks in Process Fault Diagnosis, IEEE Transactions on Systems, Man, and Cybernetics, Vol 21, No. 4, 1991.

[3] Liu, T. I. and Mengel, J. M., Intelligent Monitoring of Ball Bearing Conditions, Mechanical Systems and Signal Processing, Vol. 6, No. 5, pp 419-431, 1992.

[4] McCormick A. C., Nandi A. K., and Jack L. B., Digital Signal Processing Algorithms in Condition Monitoring, International Journal of COMADEM, vol 1., no. 3, pp 5-14, 1998.

[5] Nikias C. L. and Mendel J. M., Signal Processing with Higher Order Spectra, IEEE Signal Processing Magazine, pp 10-37, July 1993.

[6] Goldberg G. E., Genetic Algorithms in Search, Optimisation and Machine Learning, Addison Wesley, New York, 1989.