ISSN: 2277-3754
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 1, Issue 4, April 2012

Feature Selection by Genetic Programming, And Artificial Neural Network-based Machine Condition Monitoring

Pooja Gupta, Dr. Sulochana Wadhwani

Abstract*—*This Paper presents the performance of bearing fault diagnosis using Genetic Programming and artificial neural networks (ANNs). The experimental data is collected for four bearings conditions namely: Healthy, defective Outer race, defective Inner race and defective ball fault condition. Artificial neural network have been widely used for health diagnosis of rotating machinery using features extracted from vibration emission signals. One of the most important considerations in applying neural networks to condition monitoring of electrical machine is the proper selection of training features. Irrelevant or noisy features unnecessarily increase the complexity of the problem and can degrade modeling performance. A Genetic programming for feature selection is developed, based on the concept of dominance. GP is used for two purpose Feature extractor and feature selector but in this paper GP is used only for best feature selection from a large features data set. The algorithm is used to effectively select a smaller subset of features that together form a genetically fit family for fault identification and classification tasks.

***Index Terms*— Bearing Fault detection, Time domain features, genetic programming, Neural Networks.**

I. INTRODUCTION

Condition monitoring of machinery has increased in importance as more engineering processes are automated and the manpower required to operate and supervise plants is reduced. The monitoring of the condition of machinery can significantly reduce the cost of maintenance. Firstly, it can allow an early detection of potential catastrophic fault, which could be extremely expensive to repair. Secondly, it allows the implementation of conditions based maintenance rather than periodic or failure based maintenance. In this case, significant savings can be made by delaying schedule maintenance until convenient or necessary. Bearing is an essential component of any electrical motor. The function of different types of bearing is to provide slipping of the rotor inside the stator maintaining uniform air gap. The bearing consists mainly of the outer-race; inner-race, balls and the cage which assures equidistance between the balls .Bearing faults can take place due to fatigue even under normal balanced operation with good shaft alignment and can also be caused by improper lubrication, installation errors and contamination. One of the results of bearings failures are increased level of vibration and noise. In single point defect are localized and can be classified according to the affected element, outer raceways defect, inner race ways defect, ball defect and cage defect.

Neural networks have a proven ability in the area of nonlinear pattern classification. After being trained, they contain expert knowledge and can correctly identify the different causes of bearing vibration. The capacity of artificial neural networks to mimic and automate human expertise is what makes them ideally suited for handling nonlinear systems. Neural networks are able to learn expert knowledge by being trained using a representative set of data. At the beginning of a neural network’s training session, the neural network fault detector’s diagnosis of the motor’s condition will not be accurate. An error quantity is measured and used to adjust the neural network’s internal parameters in order to produce a more accurate output. This process is repeated until a suitable error is achieved. Once the network is sufficiently trained and the parameters have been saved, the neural network contains all the necessary knowledge to perform the fault detection. The appropriate pre-processing of the measurement data enables the exclusion of the data, which are less correlated to the bearing condition. Consequently, the minimization of the training vector, and thus reduction of NN-training time and computation cost can be obtained. Vibration signals have been used as input features to the ANN.

One of the most important aspects of achieving good neural network performance has proven to be the proper selection of training features. Thus feature selection which is a process of identifying those features that contribute most to the discrimination ability of the neural network is required. The aim of this paper is therefore to examine the use of Genetic programming to select the most significant input features from a large set of possible features in machine condition monitoring contexts. The results show the effectiveness of the selected features from the acquired raw and pre-processed signals in diagnosis of machine condition. This paper consists of the following tasks:

- ☐ Calculate time domain features from vibration signals using mat-lab.
- ☐ Implement a feature selection algorithm using genetic programming to minimize the number of selected features and to maximize the performance of the neural network.
- ☐ Train the neural network using a back propagation

ISSN: 2277-3754
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 1, Issue 4, April 2012

algorithm with the reduced set of features from genetic programming.

□ Investigate the effect of increasing the number of hidden nodes in the performance of the computational intelligence engine.

II. STATISTICAL FEATURES

Time domain features were extracted from the statistical measures of Median, Root mean Square (RMS), Crest factor (CRF), histogram lower bound (LB), histogram upper bound (UB), Entropy (ENT), Skewness (SK), Kurtosis (KT), Variance (VAR), Shape factor(SHF), Impulse factor(IMF), Clearance factor(CLF). Table 1 shows the time domain features.

$$\mathrm{RMS} = \sqrt{\frac{\sum_{i=1}^{N} x_i}{N}}$$

$$\mathrm{CRF} = \frac{\mathrm{Peak}}{\mathrm{RMS}}$$

$$\mathrm{LB} = \min(x) - \frac{1}{2}\frac{\max(x)-\min(x)}{N-1}$$

$$\mathrm{UB} = \max(x) + \frac{1}{2}\frac{\max(x)-\min(x)}{N-1}$$

$$\mathrm{ENT} = H(p) = -\sum_{i=1}^{N} p_i \log p_i$$

$$\mathrm{SK} = \frac{\frac{1}{N}\sum_{i=1}^{N} x_i^3}{\sigma^3}$$

$$\text{Where } \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\left(x_i - \mathrm{mean}(x)\right)\right)^2}$$

$$\mathrm{KT} = \frac{\frac{1}{N}\sum_{i=1}^{N} x_i^4}{\sigma^4}$$

$$\mathrm{VAR} = S^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}$$

$$\mathrm{SHF} = \frac{\mathrm{RMS\ value}}{\frac{1}{N}\sum_{i=1}^{N}|X_i|}$$

$$\mathrm{IMF} = \frac{\mathrm{Peak\ value}}{\frac{1}{N}\sum_{i=1}^{N}|X_i|}$$

$$\mathrm{CLF} = \frac{\mathrm{Peak\ value}}{\left(\frac{1}{N}\sum_{i=1}^{N}\sqrt{|X_i|}\right)^2}$$

Table 1. Time domain Feature extraction using vibration signals

S.N.	Extracted Features	H	DB	DIR	DOR
1.	Median	1.877	1.838	1.179	0.783
2.	RMS	0.071	0.175	1.177	0.923
3.	CRF	3.973	3.504	1.783	2.499
4.	LB	1.548	0.945	-2.803	-2.678
5.	UB	0.283	0.613	2.099	2.308
6.	ENT	3.532	4.259	3.844	4.104
7.	SK	-0.013	-0.193	-0.533	-0.014
8.	KT	2.929	3.093	2.605	2.744
9.	VAR	0.005	0.036	1.386	0.853
10.	SHF	0.037	0.095	1.129	1.148
11.	IMF	4.961	4.393	2.145	3.095
12.	CLF	5.850	5.181	2.460	3.633

III. GENETIC PROGRAMMING

Genetic programming based feature selection was used to improve the classification results and reduce the dimensionality of the data. In this paper, GP, as a form of evolutionary algorithm and an extension of genetic algorithms, is used as the feature selection. The major difference between the GP and GA approaches lies in the way that each algorithm solves the problem under consideration. With a GA-based solution, the basic form of the solution is predefined; the GA is able to optimize parameters of the solution, however not the actual structure of the solution. GP by comparison has control over both the structure and the parameters of the solution to the problem. Fig. 1 illustrates the system proposed in this paper. The block with a bold frame is the feature selector, which extracts the information from the raw vibration data to create features, based on the evolutionary algorithm. The surviving features from the feature selector are used as the inputs to the multilayer perceptron (MLP) for the classification of the four bearing conditions. Of course, some other classifier can be used as an alternative.

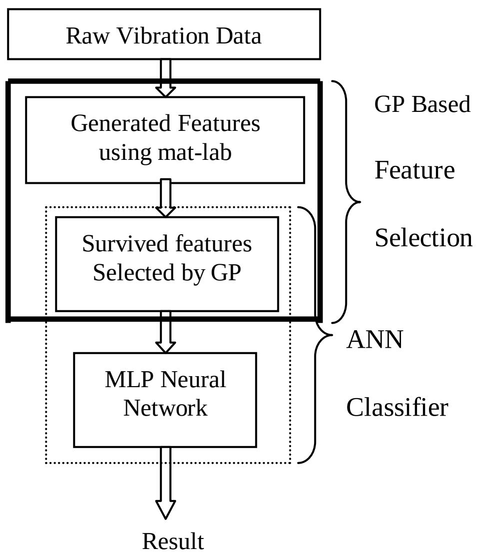

Fig. 1. Structure of the system

A. *Process of GP:*

Since the aim of bearing fault classification is to identify different machine conditions from raw vibration signals accurately, the GP-based feature selector is used to select the most significant input features for the classifiers from a large set of possible features in machine condition monitoring contexts. The purpose of GP is to try to maximize the extra information content in the sample of the raw vibration signal, and it implicitly maximizes the separation between different conditions within the data. First, an initial population with a

ISSN: 2277-3754
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 1, Issue 4, April 2012

chosen number of individuals is generated on a random basis, meaning that there is no human influence, or bias, in the selection of original features. Calculated features data set is fed as the inputs to the initial population. Each individual represents a transformation network, which tries to transform raw data into information for classification. In terms of the usefulness of each individual for classification, a fitness value is assigned to each individual by fitness function. Therefore, the members with the best fitness values survive from the current generation and will be chosen as the origins of the next generation. In our design, only the elite will survive the natural selection. This mechanism allows the feature to evolve in a direction toward the best classification performance, thus achieving the best selection of features. At the beginning of the next generation, three operations—reproduction, crossover, and mutation—are conducted to produce new members based on the surviving member. If the termination criterion is met, the best solution is preserved.

As one of the most important components of GP, the fitness function can greatly affect the performance of the system. A good fitness measure guarantees the improvement of solutions by rating the performance of each member and giving the stronger one a better chance of surviving. Traditionally, the classification results are used as the fitness value for multicategory classification problem; however, the computational demands are relatively high in training and validating a classifier for each individual.

B. Primitive Operations:

In this paper, GP as an evolutionary method is proposed to minimize the correlation coefficient of the extracted features data set. Table 2 represents the correlation coefficient of twelve features used in GP. GP evolves tree individuals representing possible solutions to the problem at hand. A population of such individuals is randomly created and then evolved by probability of genetic operations:

• Crossover: GP carries out a crossover operation to create new individuals with a probability, which controls the occurrence of the crossover throughout generations. Two new individuals are generated by selecting compatible nodes randomly from each parent and swapping them.

• Mutation: The mutation operation is performed by the creation of a sub-tree at a randomly selected node with the probability. First, for a given parent, there is an index assigned to each node for identification. A random index number is generated to indicate the place where mutation will happen. The node is located, then the tree downstream from this node is deleted and a new sub tree is generated from this node, exactly in the same way as growing initial population.

• Reproduction: The reproduction operation is performed by copying individuals to the next population without any change in terms of a certain probability. All these three operations happen within one generation based on three probabilities.

C. Primitive Terminators:

Terminators act as the interface between GP and the correlation coefficient of the features data set. They are required to collect fault-related information as much as possible from the feature data set and to provide inputs to the feature selector. In our GP-based feature selector, the terminator set is constructed by computing the estimate of twelve statistical moments.

IV. CLASSIFIERS

There are three possible ways of using classifiers.

• In the first case, raw vibration data can be used directly as the inputs of the classifier. We leave it to the classifiers for selection and extraction of information from the raw data by enhancing discriminating features and diminishing interfering data during the learning process. This is carried out by adjusting weights and the amount of data processed by the classifier will be enormous. Therefore, the useful information is difficult to maximize, while interference cannot be eliminated entirely.

• In the second case, a much more popular and effective realization, some useful features are prepared from the raw data for the inputs to the classifier. Based on the physical and mathematical analysis of the mechanism of rotating machinery in faulty conditions, features are computed to represent the fault information concealed within the raw vibration data. Therefore, classifier has much less difficulty than in the first realization.

• In the third case, rather than manual development, features are selected through an evolutionary process in order to avoid human negative influence and bias, and to conduct the feature extraction from a much larger space.

Obviously, the latter two realizations are more effective and are examined and compared through a series of experiments in this paper. To demonstrate the robustness of extracted features, classification algorithms are proposed, including artificial neural networks (ANNs).

V. ANN

ANNs are probably one of the most common classifiers in use today. This is mainly due to their ability to learn and identify patterns in the source data. For machine condition monitoring, where the training dataset is often sparse, and the classifier has to generalize to a certain extent, ANN is an ideal solution because of its nonlinearity, and many applications can be found regarding bearing fault detection. The multilayer peceptron (MLP) is chosen here as the structure of the network for its overall performance over other configurations. The MLP used consists of one hidden layer and one output layer, with the hidden layer having a logistic activation function and the output layer using a linear activation function. For training procedure, the back propagation algorithm with adaptive learning and momentum is used. The learning algorithm is stopped when the classification performance of the validation set starts to

ISSN: 2277-3754
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 1, Issue 4, April 2012

diverge from that of the training set.

Table 2. Correlation Coefficient of twelve features used in GP

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
F1	1	-0.803	-0.877	0.777	0.128	0.806	-0.956	0.803	0.798	-0.265	-0.969	0.943
F2	-0.803	1	0.988	-0.943	-0.665	-0.984	0.942	-0.986	-0.986	0.042	0.909	-0.949
F3	-0.877	0.988	1	-0.918	-0.580	-0.986	0.978	-0.987	-0.986	0.149	0.961	-0.985
F4	0.777	-0.943	-0.918	1	0.502	0.870	-0.895	0.876	0.877	0.251	-0.837	0.878
F5	0.128	-0.665	-0.580	0.502	1	0.690	-0.402	0.694	0.699	-0.049	-0.364	0.449
F6	0.806	-0.984	-0.986	0.870	0.690	1	-0.938	0.999	0.999	-0.216	-0.924	0.956
F7	-0.956	0.942	0.978	-0.895	-0.402	-0.938	1	-0.937	-0.935	0.182	0.993	-0.997
F8	0.803	-0.986	-0.987	0.876	0.694	0.999	-0.937	1	0.999	-0.202	-0.922	0.955
F9	0.798	-0.986	-0.986	0.877	0.699	0.999	-0.935	0.999	1	-0.196	-0.919	0.953
F10	-0.265	0.042	0.149	0.251	-0.049	-0.216	0.182	-0.202	-0.196	1	0.291	-0.234
F11	-0.969	0.909	0.961	-0.837	-0.364	-0.924	0.993	-0.922	-0.919	0.291	1	-0.994
F12	0.943	-0.949	-0.985	0.878	0.449	0.956	-0.997	0.955	0.953	-0.234	-0.994	1

VI. RESULTS AND COMPARISON

Tables 3 are achieved from combinations of each of 2 to 10 GP based-selected features and an MLP with one hidden layer of each of 10 neurons. Table 4 shows percentage of the Comparison study of classification success with MLP, which consists of one hidden layer with 10 neurons, Using Different Number of features without GP selected Features/ANN and using GP based selected features from 2 to 10. It is clear that the GP/ANN classification success rate is always more than 99% and without GP/ANN classification success rate obtained minimum 78.42 for 2 features and maximum 96.48 for 10 features.

It can be seen that using the GP selected features the classification results are robust with respect to the choice of the number of neurons. Classification performances improve by increasing the number of neurons and increasing the number of features.

The computation cost of GP in the feature selection process is slightly larger. However, GP requires less computation compared with GA for feature selection and generation. GP also has the ability to process the un-normalized datasets as the input, remove the systematic variation and bring the raw data onto the same ground for a fair comparison.

Table 3 Classification Result Using Different Number of GP based selected Features

S. N	No. Of features Selected by GP	Features name	GP/ANN Accuracy (%)	Max. Fitness Calculated by GP
1.	2	Median, SK	99.86%	0.6067
2.	3	Median, RMS, SK	99.88%	0.4779
3.	4	Median, RMS, SK, SHF	99.91%	0.4363
4.	5	Median, RMS, SK, SHF, ENT	99.92%	0.4225
5.	6	Median, RMS, SK, SHF, ENT,UB	99.95%	0.4001
6.	7	Median, VAR, RMS, SK, SHF, CLF, ENT	99.97%	0.3967
7.	8	Median, VAR, RMS, SK, SHF, CLF, ENT,UB	99.99%	0.3826

ISSN: 2277-3754
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 1, Issue 4, April 2012

8.	9	Median, VAR, RMS, SK, CRF ,SHF, IMF , CLF, ENT	100%	0.3735
9.	10	Median, VAR, RMS, SK, CRF ,SHF, IMF , CLF, ENT,UB,	100%	0.3659

Table 4 .Comparison of Classification Result using without GP/ANN and with GP based selected Features/ANN

S. N.	No. of Features	Without GP ANN classification rate	With GP based selected features ANN classification rate
1	2	78.42%	99.86%
2	3	79.94%	99.88%
3	4	81.91%	99.91%
4	5	82.46%	99.92%
5	6	89.55%	99.95%
6	7	92.73%	99.97%
7	8	94.62%	99.99%
8	9	95.55%	100%
9	10	96.48%	100%

V. CONCLUSION

In this paper, a GP-based feature selector is proposed for the selection of best features from the large features data set for classification applied to the problem of bearing fault classification. GP is a powerful and efficient tool for the feature selection directly from the features data set. Using features selected by GP, the ANN classifier sees a significant improvement in classification results. It is also shown from the selection results that GP is not only capable of enhancing the classification performance, but also reducing the dimensionality to describe the problem. Furthermore classification performances obtained from GP selected features are very robust. Also, GP produces results in a tree representation, which allows an understanding of how it works.

REFERENCES

[1] L. B. Jack and A. K. Nandi, "Genetic algorithms for feature selection in machine condition monitoring with vibration signals," Proc. Instit. Elec. Eng. Vision, Image Signal Processing, vol. 147, no. 3, pp. 205–212, Jun. 2000.

[2] G. A. Rovithakis, M. Maniadakis, and M. Zervakis, "A hybrid neural network/genetic algorithm approach to optimizing feature extraction for signal classification," IEEE Trans. Syst., Man, Cybern. Part.B, vol. 34, no. 1, pp. 695–703, Feb. 2004.

[3] M. Brameier and W. Banzhaf, "A comparison of linear genetic programming and neural networks in medical data mining," IEEE Trans. Evol. Comput., vol. 5, no. 1, pp. 17–26, Feb. 2001.

[4] J. K. Kishore, L. M. Patnaik, V. Mani, and V. K. Arawal, "Application of genetic programming for multicategory pattern classification," IEEE Trans. Evol. Comput., vol. 4, no. 3, pp. 242–258, Sep. 2000.

[5] L. Zhang, L. B. Jack, and A. K. Nandi, "Fault detection using genetic progamming," Mech. Syst. Sig. Process., 2003.

[6] J. R. Sherrah, R. E. Bogner, and A. Bouzerdoum, "The evolutionary pre-processor: Automatic feature extraction for supervised classification using genetic programming," in Proc. 2nd Int. Conf. Genetic Programming, 1997, pp. 304–312.

[7] M. Kotani, S. Ozawa, M. Nasak, and K. Akazawa, "Emergence of feature extraction function using genetic programming," in Proc. 3rd Int. Conf. Knowledge-Based Intelligent Information Engineering Systems, 1997, pp. 149–152.

[8] Hong Guo, Lindsay B. Jack, and Asoke K. Nandi, "Feature Generation Using Genetic Programming With Application to Fault Classification" IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 35, NO. 1, FEBRUARY 2005.

[9] J. Yang and V. Honvar, "Feature subset selection using a genetic algorithm," IEEE Intell. Syst., vol. 13, pp. 44–49, Mar./Apr. 1998.

[10] M. L. Raymer, W. F. Punch, E. D. Goodman, L. A. Kuhn, and A. K. Jain, "Dimensionality reduction using genetic algorithms," IEEE Trans. Evol. Comput., vol. 4, pp. 164–171, Apr. 2000.

[11] J. R. Koza, Genetic Programming: On the Programming of Computers by Means of Natural Selection. Cambridge, MA: MIT Press, 1992.

[12] I. Benyahia and J. Potvin, "Decision support for vehicle dispatching using genetic programming," IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, vol. 28, no. 3, pp. 306–314, May 1998.

[13] M. J. Chang and A. K. Jain, "Artificial neural networks for feature extraction and multivariate data projection," IEEE Trans, Neural Netw. vol. 6, no. 2, pp. 296–317, Mar. 1995.

[14] P. Chen, T. Toyota, and Z. He, "Automated function generation of symptom parameters and application to fault diagnosis of machinery under variable operating conditions," IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, vol. 31, no. 6, pp. 775–781, Nov. 2001.

[15] A. C. McCormick and A. K. Nandi, "Real time classification of rotating shaft loading conditions using artificial neural networks," IEEE Trans. Neural Netw., vol. 8, no. 3, pp. 748–757, May 1997.

AUTHOR BIOGRAPHY

Pooja Gupta received the B.E. degree in Electrical Engineering from MITS Gwalior, in 2009 and is currently pursuing ME degree in Specialization of Measurement and control from the same institute. Her area of interests is in field machine and control system.

Dr. S. Wadhwani is the reader in Electrical Engineering Department, Faculty of Engineering; MITS Gwalior (M.P.). Her area of research interest includes control system, drives & Instrumentation.