



9th International Conference on Digital Enterprise Technology - DET 2016 – “Intelligent Manufacturing in the Knowledge Economy Era

Dynamic Genetic Algorithm-based Feature Selection Scheme for Machine Health Prognostics

Lei Lu, Jihong Yan*, Yue Meng

School of Mechatronics Engineering, Harbin Institute of Technology, Harbin 150001, China

* Corresponding author. Tel.: +86 0451 86402972; fax: +86 0451 86402972. E-mail address: jyan@hit.edu.cn

Abstract

This paper proposed a dominant feature selection scheme to enable the high performance prognostics of machine health. Statistical features were extracted from decomposed sub-modes by wavelet transform. Fisher ratio was employed to evaluate the extracted feature vectors, and dynamic searching strategy-based genetic algorithm was used to select the optimal feature subsets on the basis of maximizing the fitness function. Then dominant features with minimum mean square errors were used to predict the performance of machine health. Experimental results on predicting the lifetime of an unbalance vibration rotor system demonstrated that the proposed method can achieve better prognosis performance with less predicting errors.

© 2016 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the scientific committee of the 5th CIRP Global Web Conference Research and Innovation for Future Production

Keywords: Machine health prognostics; feature selection; genetic algorithm; wavelet transform.

1. Introduction

Rotary mechanical facilities are widespread in almost every aspect of industries. Machine health degradation or unexpected failures may impact product quality, lead to significant economic losses, and even cause potential safety hazard. Prognostics are regarded as significant and promising tools for achieving benefits of reducing the equipment downtime, maintenance cost, and safety accidents [1] by utilizing on-line information from machine health monitoring and the off-line priori knowledge originated from expert system.

Intelligent prognostic system has attracted much attention in the fields of condition-based maintenance during the past decades. Statistical methods such as auto regressive moving average model and support vector machine have been employed for machine health assessment and residual useful life (RUL) prediction [2]. A combination technique of relevance vector machine-based data driven-approach and exponential regression-based model has been developed to

estimate the RUL of ball bearings and analysis vibration-based degradation data [3]. A robust online prognostic system is developed by integrating collaborative physics-of-failure modelling, reliability engineering, damage accumulation modelling [4]. Knowledge discovery techniques-based prognostic model is presented for the accurate assessment of the probability of flight-critical hardware failure [5].

Generally, prognostic system consists of signal processing, feature selection, performance evaluation, etc. Sensor information of the current machine condition is used to extract meaningful features, however, a rich set of signal processing techniques such as spectrum analysis, probabilistic approaches, and time-series analysis are employed to generate feature candidates, which result in the challenging task of identifying dominant features from high dimensionality feature space. Irrelevant or noisy features may increase complexity of the prognostic model, degrade the evaluation performance and reduce the learning speed. Moreover, it is difficult to determine key components that decomposed by time-frequency methods such as wavelet analysis (WT) or

empirical mode decomposition. Some literatures only selected several decomposed sub-modes or empirically extracted fault features [6].

Research has been carried out into feature selection (FS) techniques over the past several decades [7-9]. The search for the optimal subset is essentially an NP-hard problem, especially in a high-dimension feature space. Obtaining of the optimal solution cannot be guaranteed without performing an exhaustive search in the solution space, and this technique is impractical for large scale features because of the unaffordable time cost [10]. The genetic algorithm, a form of inductive learning strategy, has great advantage for efficient feature selection, which provides near-optimal solutions in complex and nonlinear search spaces with time efficiency [7, 8]. However, the traditional genetic algorithm (GA) with fixed length chromosomes and stationary searching space of gene bases cannot be directly employed to find the optimal feature sub-sets. Moreover, to achieve the global optimum result, interactive relationships between features within the same chromosome need to be developed [11].

In this paper, a dominant feature selection scheme for rotary machinery health prognostics is developed based on an innovative variable-range genetic algorithm (VRGA), which employs a dynamic searching strategy by varying the length of chromosomes and changing in real-time the range of the feature candidates. Statistical features are extracted from sub-modes decomposed by WT and dominant features are optimized through the developed VRGA model. Then the selected dominant features are employed to predict the lifetime of a rotary mechanical system.

2. Wavelet transform

The continuous wavelet transform (CWT) of time series $x(t)$ is defined as a dot product of signal $x(t)$ and the wavelet basis function [12],

$$cwt(a, b) = \langle x(t), \psi_{a,b}(t) \rangle = \frac{1}{\sqrt{|a|}} \int x(t) \psi\left(\frac{t-b}{a}\right) dt \quad (1)$$

where, $\psi_{a,b}$ is the wavelet basis, parameters a and b are used to scale and translate the mother wavelet function $\psi(t)$, and $cwt(a, b)$ is the coefficient of continuous wavelet transform.

During the process of calculating wavelet coefficients in CWT, the scale parameter a and translation factor b are continuously changing, which may result in intractable computations and generate an awful lot of data [13]. Discrete wavelet transform (DWT) provides an efficient solution by employing discrete scaling and translation factors (i.e., $a = 2^j$, $b = k2^j$), then, the wavelet transform of signal $x(t)$ can be expressed as,

$$dwt(j, k) = \frac{1}{\sqrt{2^j}} \int x(t) \psi\left(\frac{t-k2^j}{2^j}\right) dt \quad (2)$$

The operation of DWT can be realized on the basis of multi-resolution analysis by employing a half band low-pass filter $h(k)$ and a half band high-pass filter $g(k)$, which are derived from the mother wavelet function $\psi(t)$ and the corresponding scaling function $\phi(t)$ [14, 15]. The original signal is decomposed into a set of approximations and detailed components,

$$a_{j,k} = \sum_k h(n-2k)a_{j-1,k} \quad (3)$$

$$d_{j,k} = \sum_k g(n-2k)a_{j-1,k} \quad (4)$$

The original signal $x(t)$ is decomposed as,

$$x(t) = \sum_k a_{j,k} 2^{-j/2} \phi(2^{-j}t - k) + \sum_k d_{j,k} 2^{-j/2} \psi(2^{-j}t - k) \quad (5)$$

where, j is the decomposition level, k is the discrete time, $a_{j,k}$ represents the low-frequency approximations of the original signal, and $d_{j,k}$ corresponds to the higher frequencies.

Table 1. Feature and nomenclature.

No.	Feature name	Notation
1	Mean value	F _M
2	Maximum point	F _{Max}
3	Minimum value	F _{Min}
4	Standard deviation	F _{SD}
5	Peak-peak value	F _{Pp}
6	Kurtosis	F _{KR}
7	Pulse index	F _P
8	Waveform factor	F _{WF}
9	Energy of each IMF	F _E
10	The 5 th central moment	F ₅
11	Shannon entropy	F _{EP}

Statistical features shown in Table 1 are extracted from each decomposed level. Shannon entropy is calculated from all sub-bands. Intuitively, it is difficult to estimate which features are sensitive to describe characteristics of the investigated signal, as various factors result in the mechanical failure. Here, 11 features which are widely used are employed and there is no need for the laborious tasks of signal filtering, framing, windowing, and so on [16, 17].

3. The proposed feature selection scheme

The present paper develops a feature selection scheme by employing a novel two-dimensional searching genetic algorithm. Flow chart of the proposed machine health prognostic scheme is shown in Figure 1. The investigated signal is decomposed into sub-modes by the WT method after signal pre-processing. Then the developed dynamic searching algorithm is used to find the optimized feature subset of the extracted statistical features, and dominant features are selected with minimum mean square error. Finally, time series predicting model is used to evaluate the current performance of machine health.

3.1. Design of the structure of chromosomes

In order to search through the space of the set of feature space and encode all feature candidates within one chromosome, a feature representation solution is developed. Given a set of N -dimensional feature pool, an M -dimensional feature subset ($M \leq N$) is initialized, as shown in Figure 2. Each feature candidate in one of the total P -chromosomes has a precision n . There are three relationships in this solution: (1)

all feature candidates must differ from each other within the same chromosome, (2) dimensionality of the feature subset grows with the number of iterations, and (3) the range of feature candidates from #1 to #M decreases in each chromosome.

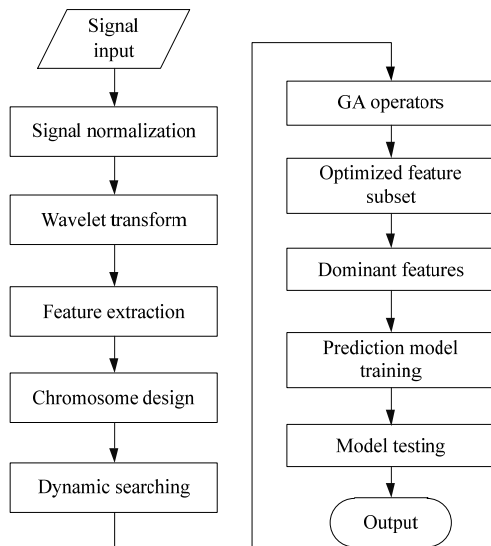


Fig. 1. Flow chart of the proposed machine health prognostic system.

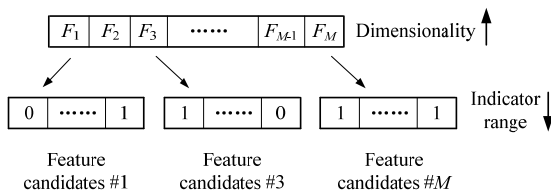


Fig. 2. Chromosome structure design.

3.2. Defining the fitness function

The chromosomes are transformed into real values and evaluated by the fitness function. The Fisher criterion is a method widely used in statistics and machine learning to find the linear combination of features which best separates two or more classes of datasets [18], which is defined as follows,

$$J(d) = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (\mu_i - \mu_j)^2}{\sum_{i=1}^n \sigma_i^2} \quad (6)$$

Where, μ_i is the mean vector, σ_i is the variance matrix of class i , n is the total number of faults, and $i, j = 1, 2, 3, \dots, n$.

Fisher criterion finds the linear combination of features by maximizing the ratio, which indicates that different stages have the maximum within-class distance and minimum between-class value. Then the corresponding feature subset is regarded as the optimal one.

3.3. The GA optimization process

The process of GA optimization [19] is to maximize the Fisher ratio of different training sample. First, for the selection operation, all chromosomes are ranked on the basis of the corresponding fitness values, and elite chromosomes are selected as parent chromosomes according to the pre-set selection rate, which are participate in producing offspring for the next generation. Then different parent chromosomes change genes to create new chromosomes through the single-point crossover process. Finally, the mutation process is carried out, a small portion of individuals randomly change one or more bits to create child chromosomes.

4. Time series prediction

In our application, the unbalance vibration is in proportion to the eccentric mass, which is simulated by gradually increasing standard weights to the wheel. The selected dominant features are used to predict machine health based on the time series forecasting method. Sensor information that represents the current condition of machine health is used to extracted statistical features.

Let $f_{m,k} \in \mathbb{R}^m$ donates an m dimensional feature vector of the k stage of machine health, which is employed to forecast the performance of the current condition. The standard prediction function takes the form as [17],

$$\begin{bmatrix} f_{m,1}^T \\ f_{m,2}^T \\ \vdots \\ f_{m,n}^T \end{bmatrix} \times \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_m \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix} \quad (7)$$

Where, θ_i is the prediction coefficient that is trained by the historical data, d_k is the desired performance of machine health.

Using the proposed feature selection method, p dominant features are selected. The corresponding feature vector and prediction coefficient are expressed as,

$$F = \begin{bmatrix} f_{p,1}^T \\ f_{p,2}^T \\ \vdots \\ f_{p,n}^T \end{bmatrix}, \varphi = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{bmatrix} \quad (8)$$

Then, the performance of stage k can be calculated as,

$$d_k = f_{p,k}^T \theta \quad (9)$$

For n times forecasting, the prediction error is computed as,

$$MSE = \frac{1}{n} \sum_{k=1}^n (D_k - f_{p,k}^T \theta)^2 \quad (10)$$

where, D_k is the real performance of the k stage.

5. Experiments and discussion

The experiment is conducted on a rotor degradation

system shown as Figure 3. The test-bed consists of a three-phase induction motor (left), two eddy current sensors (centre), a rotor and control electronics. There are sixteen screw holes uniformly distributed on the face of the mass wheel in order to add standard weights. Performance degradation of the test rig is simulated by gradually increasing the number of standard weights. The rotating speed is 2400rpm and vibration signals are probed with sampling frequency 1,000 Hz. Each stage of the experiments has 40 experiment samples, 24 samples are randomly selected to train fault diagnosis model, and the rest are used for model testing.

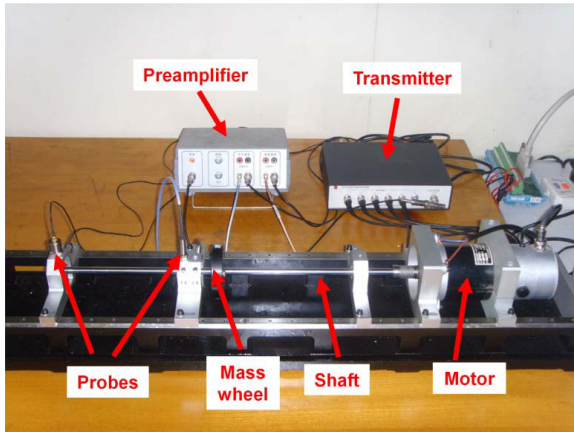


Fig. 3. Rotor degradation system.

Parameters of the wavelet transform are carefully selected, and statistical features discussed in Section 2 are calculated from all the decomposed sub-bands. Statistical features of the wavelet sub-modes are shown as Figure 4, the vibration signal is decomposed into 6 levels using db4 wavelet. It should be noted that the previous 10 features listed in Table 1 are calculated from each of the 7 sub-modes, while Shannon entropy is extracted based on all 7 sub-modes. Thus, a feature vector with 71 dimensionality of each sample is obtained.

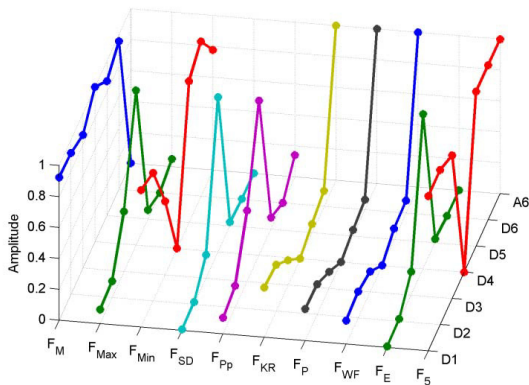


Fig. 4. Statistical feature of WT sub-modes.

The proposed method selects the optimal feature subsets with dimensionality increasing from 1 to 71, and sends them to the time series forecasting model. Parameters of the employed GA are carefully selected, with 400 iterations and 50 individuals. The precision of each variable is 20, the generation gap is 0.9, recombination rate is 0.7, probability of mutation is 0.7, and the roulette strategy is employed. As shown in Figure 5, with generations increasing, the Fisher ratio tends to be constant, while this ratio is decreasing with the feature dimensionality growing. Distribution of the optimized feature vector with 18 dominant features is shown in Table 2, where the symbol ‘+’ denoted features that are selected. It can be seen that these dominant features are mainly distributed in the 3rd, 4th and 5th detail sub-bands that are decomposed by WT.

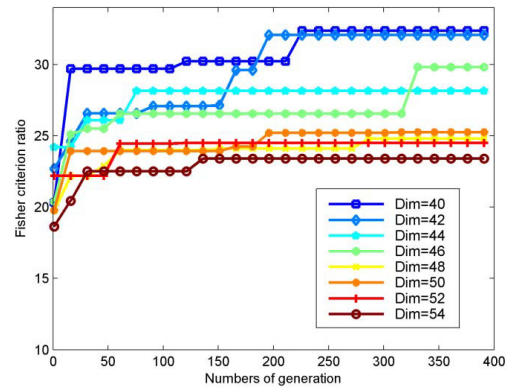


Fig. 5. Training of the VRGA model.

Table 2. Dominant features distribution.

Sub-mode	F _M	F _{Max}	F _{Min}	F _{SD}	F _{Pp}	F _{KR}	F _P	F _{WF}	F _E	F ₅	F _{Ep}
A6											
D6											
D5			+	+							+
D4		+	+	+	+	+	+	+			+
D3		+	+	+	+						
D2					+						
D1			+								+

The optimized feature subsets with feature dimensionality ranging from 1 to 71 are used to predict the lifetime of the machine health. Figure 6 shows the comparison of prediction results with 8 features, 18 features and all features. The prediction performance is evaluated by mean square error (MSE), mean relative error (MRE), and mean absolute error (MAE). As shown in Table 3, the prediction performance receives the minimum error when the feature dimensionality is 18, while the maximum error occurs with only 8 features.

Table 3. Prediction performance evaluation.

Feature dimensionality	MSE	MRE	MAE
8	0.2627	0.9758	0.9733
18	0.0738	0.1592	0.0719
30	0.1051	0.1805	0.1273
45	0.1169	0.2104	0.2150
60	0.1259	0.4119	0.5652
71	0.1509	0.4925	0.6762

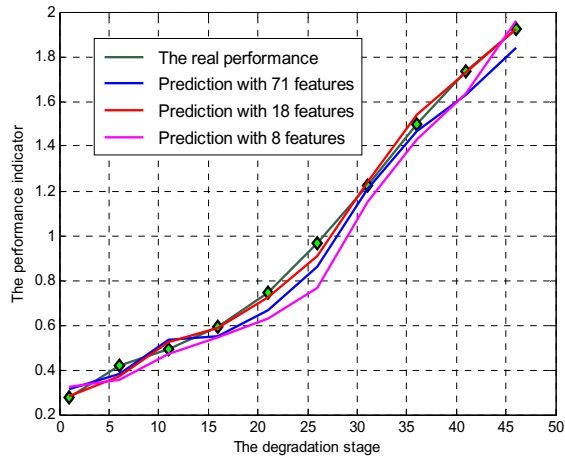


Fig. 6. Machine health prognostics

6. Conclusion

This paper developed a dynamic searching strategy-based genetic algorithm for machine health prognostics. Results on analysing performance degradation of an unbalance vibration rotor system demonstrated that the proposed method can effectively predict the current condition of machine health. Meanwhile, the optimized feature subsets indicated that the 3rd, 4th and 5th detail sub-bands decomposed by WT contain meaningful information for mechanical prognostics. Future research will mainly focus on identifying intrinsic manifolds from machine health motoring information, which may assist better understanding of the variation for health condition.

Acknowledgements

This research is funded by the National Natural Science Foundation of China (#51561125002).

References

- [1] Lu C, Tao L, and Fan H, An intelligent approach to machine component health prognostics by utilizing only truncated histories. *Mechanical Systems and Signal Processing*, 2014. 42(1-2): p. 300-313.
- [2] Widodo A and Yang B-S, Machine health prognostics using survival probability and support vector machine. *Expert Systems with Applications*, 2011. 38(7): p. 8430-8437.
- [3] Di Maio F, Tsui KL, and Zio E, Combining Relevance Vector Machines and exponential regression for bearing residual life estimation. *Mechanical Systems and Signal Processing*, 2012. 31: p. 405-427.
- [4] Baybutt M, et al., Improving Digital System Diagnostics Through Prognostic and Health Management (PHM) Technology. *Instrumentation and Measurement, IEEE Transactions on*, 2009. 58(2): p. 255-262.
- [5] Skormin VA, Gorodetski VI, and Popyack LJ, Data mining technology for failure prognostic of avionics. *Aerospace and Electronic Systems, IEEE Transactions on*, 2002. 38(2): p. 388-403.
- [6] Ruoyu L and He D, Rotational Machine Health Monitoring and Fault Detection Using EMD-Based Acoustic Emission Feature Quantification. *IEEE Transactions on Instrumentation and Measurement*, 2012. 61(4): p. 990-1001.
- [7] Behbahani S and de Silva CW, Niching Genetic Scheme With Bond Graphs for Topology and Parameter Optimization of a Mechatronic System. *Mechatronics, IEEE/ASME Transactions on*, 2014. 19(1): p. 269-277.
- [8] Gamage LB, de Silva CW, and Campos R, Design evolution of mechatronic systems through modeling, on-line monitoring, and evolutionary optimization. *Mechatronics*, 2012. 22(1): p. 83-94.
- [9] Maldonado S, Weber R, and Basak J, Simultaneous feature selection and classification using kernel-penalized support vector machines. *Information Sciences*, 2011. 181(1): p. 115-128.
- [10] Yusta SC, Different metaheuristic strategies to solve the feature selection problem. *Pattern Recognition Letters*, 2009. 30(5): p. 525-534.
- [11] Lu L, Yan J, and de Silva CW, Dominant feature selection for the fault diagnosis of rotary machines using modified genetic algorithm and empirical mode decomposition. *Journal of Sound and Vibration*, 2015. 344(0): p. 464-483.
- [12] Sung CK, Tai HM, and Chen CW, Locating defects of a gear system by the technique of wavelet transform. *Mechanism and Machine Theory*, 2000. 35(8): p. 1169-1182.
- [13] Eldin AAH and Refaey MA, A novel algorithm for discrimination between inrush current and internal faults in power transformer differential protection based on discrete wavelet transform. *Electric Power Systems Research*, 2011. 81(1): p. 19-24.
- [14] Mallat SG, A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 1989. 11(7): p. 674-693.
- [15] Barakat M, et al., Self adaptive growing neural network classifier for faults detection and diagnosis. *Neurocomputing*, 2011. 74(18): p. 3865-3876.
- [16] Malhi A and Gao RX, PCA-based feature selection scheme for machine defect classification. *Instrumentation and Measurement, IEEE Transactions on*, 2004. 53(6): p. 1517-1525.
- [17] Jun-Hong Z, et al., Intelligent Diagnosis and Prognosis of Tool Wear Using Dominant Feature Identification. *Industrial Informatics, IEEE Transactions on*, 2009. 5(4): p. 454-464.
- [18] Huang Y, et al., A lean model for performance assessment of machinery using second generation wavelet packet transform and Fisher criterion. *Expert Systems with Applications*, 2010. 37(5): p. 3815-3822.
- [19] Holland JH, *Adaptation in Natural and Artificial Systems*. 1975: The University of Michigan Press.