

Performance assessment and fault classification for hydraulic pump based on LMD and LR

Yu Ding¹, Jian Ma², Ye Tian³

Science and Technology on Reliability and Environmental Engineering Laboratory, Beijing, China

School of Reliability and Systems Engineering, Beihang University, Beijing, China

¹Corresponding author

E-mail: ¹iconstantine@foxmail.com, ²majian3128@126.com, ³comeontianye@126.com

(Accepted 13 September 2014)

Abstract. Real-time health monitoring and fault diagnosis system of hydraulic pump is very crucial as the pump being the power source of the entire hydraulic system. Prognostic method based on logistic regression for health assessment and fault classification is proposed. The real-time state of the system is obtained by processing the data of vibration signals collected from the pumps, and maintenance can be performed as long as the failure or malfunction prognosis indicates instead of periodic maintenance inspections. The vibration signal is decomposed into several product functions by local mean decomposition (LMD), and the product functions that contain fault information form a feature vector by abstracting energy values and corresponding time-domain statistical magnitudes. Principle component analysis (PCA) is used for feature reduction. Logistic regression (LR) models are trained by the reduced features to obtain machine health condition and classify possible failure models. The maximum likelihood method is applied to determine the parameters of LR models. The methodology has been applied to process the vibration signals of a real hydraulic pump to verify the effectiveness and feasibility.

Keywords: hydraulic pump, health assessment, fault classification, local mean decomposition, principle component analysis, logistic regression model.

1. Introduction

Hydraulic pump is critical to hydraulic system and its state always reflects whether the operation of the entire system is normal or not. Generally hydraulic pumps under an abnormal state are accompanied with vibration changes [1-3]. This paper provides a prognostic method for hydraulic pump health assessment and root cause classification by logistic regression.

Four sections are included to demonstrate the whole study. Section two illustrates the prognostic methodology along with related mathematics. Section three shows the information of hydraulic pump vibration signals and assessment/classification results obtained by processing them using the proposed scheme. Section four summarizes the paper and gives future research directions.

2. Methodology

The prognostic scheme is based on monitored data which contains hydraulic pump's incipient failure signatures together with intelligent mathematical techniques which can be incorporated to detect, evaluate the risk of failure over some protracted period of time and also classify which particular type of failure may occur.

2.1. Procedures of the methodology

There are two major steps of the methodology. Step 1: feature extraction and selection that extracts and determines appropriate features for health assessment and root cause classification as well as reducing the search space for fast computation. Step 2: health assessment and root cause classification by means of the logistic regression method. The process of the methodology can be summarized as Fig. 1.

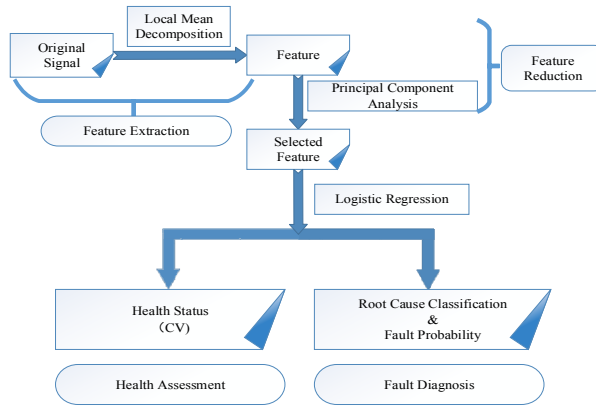


Fig. 1. Process of the methodology

2.2. Feature extraction using local mean decomposition

The LMD scheme involves separating a frequency modulated signal from an amplitude modulated envelope signal. The process of the separation includes smoothing the original signal, subtracting the smoothed signal from the original signal and using an envelope estimate to amplitude demodulate the result. Product function is obtained through multiplying the frequency modulated signal by the final envelope estimate. The whole process is repeated on the resulting signal, to acquire a second product function with an associated envelope and instantaneous frequency. By the time the remaining signal contains no more oscillations, the decomposition is finally completed [4].

The feature vector of the hydraulic pump vibration signals contains twelve components. The feature vector T is:

$$T = \left[g_{11}, g_{12}, g_{21}, g_{22}, \dots, g_{m1}, g_{m2}, \frac{E_1}{E}, \frac{E_2}{E}, \dots, \frac{E_m}{E} \right], \quad (1)$$

where the $g_{11}, g_{21}, \dots, g_{m1}$ and $g_{21}, g_{22}, \dots, g_{m2}$ are the skewness coefficients and kurtosis coefficients of the product functions, given by:

$$g_{i1} = \frac{1}{N} \sum_{i=1}^N \left[\frac{x_i - u}{\sigma} \right]^3 = \frac{E[x - u]^3}{\sigma^3}, \quad (2)$$

$$g_{i2} = \frac{1}{N} \sum_{i=1}^N \left[\frac{x_i - \mu}{\sigma} \right]^4 = \frac{E[x - \mu]^4}{\sigma^4}. \quad (3)$$

The energy value of each product function is given by:

$$E_i = \int_{-\infty}^{+\infty} |PF_i(t)|^2 dt. \quad (4)$$

In consideration of the energy values are always large, for the sake of processing and analysis, the energy vector is normalized as [5]:

$$E = \left(\sum_{i=1}^m |E_i|^2 \right)^{1/2}. \quad (5)$$

2.3. Feature reduction based on PCA

Assuming that an operation stage S (in this paper, only the normal machine behavior and machine operation with a worn piston shoe and swashplate/valve plate rotor) of the hydraulic pump is given and its features X are characterized by the multivariate Gaussian distribution with mean $\bar{\mu}_s$ and covariance matrix K_s . The symmetric matrix K_s is given as:

$$K_s = \sum_{i=1}^r \lambda_i \bar{v}_i \bar{v}_i^T = V \Lambda V^T, \tag{6}$$

where r is the rank of K_s , λ_i ($i = 1, 2, \dots, r$) are the non-zero eigenvalues of K_s and \bar{v}_i are the corresponding unit norm eigen-vectors. The definition of V and Λ are given by:

$$V = [\bar{v}_1, \bar{v}_2, \dots, \bar{v}_r], \quad \Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_1 & & \vdots \\ & & \ddots & 0 \\ 0 & \dots & 0 & \lambda_1 \end{bmatrix}. \tag{7}$$

Each eigenvalue λ_i ($i = 1, 2, \dots, r$) depicts the amount of the covariance matrix energy projected in the direction of the corresponding eigenvector. When there exists a high degree of correlation among the components of X , only a few of the eigenvalues in Λ account for most of the energy in the covariance matrix K_s . Thus, assuming that eigenvalues λ_i ($i = 1, 2, \dots, r$) are arranged in descending order, Eq. (6) can be represented as:

$$K_s = \sum_{i=1}^p \lambda_i \bar{v}_i \bar{v}_i^T = V_p \Lambda_p V_p^T, \tag{8}$$

where:

$$V = [\bar{v}_1 \quad \bar{v}_2 \quad \dots \quad \bar{v}_p], \quad \Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & & \vdots \\ & & \ddots & 0 \\ 0 & \dots & 0 & \lambda_p \end{bmatrix}, \tag{9}$$

where p is the number of the principal components of K_s , λ_i ($i = 1, 2, \dots, p$) are the largest p eigenvalues of K_s , and \bar{v}_i are the corresponding unit norm eigenvectors.

A query item \tilde{X} can now be transformed into a p component random variable \tilde{Y} given as:

$$\tilde{Y} = T(\tilde{X} - \bar{\mu}_s), \quad T = \Lambda_p^{-1/2} \Lambda_p. \tag{10}$$

If \tilde{X} belongs to the class of signals from centrifugal pump state S , then \tilde{Y} should be normally distributed with zero mean and variance I_p , where I_p is the unity matrix of order p . Thus, for each query item \tilde{X} , its adherence to the class S can be assessed through the Euclidean norm of the vector \tilde{X} , which in turn corresponds to assessment and classification based on the logistic regression of the query item from the training classes [6].

2.4. Logistic regression method

The machine condition description from daily maintenance records/logs is a dichotomous problem (either normal or failed) which can be represented using a logistic regression function [1]. The goal of logistic regression is to find the best fitting model to describe the relationship between

the categorical characteristic of dependent variable (the probability of an event, constrained between 0 and 1) and a set of independent variables. The logistic function is:

$$prob(event) = p(x) = (1 + e^{-g(x)})^{-1}. \quad (11)$$

The logistic model is:

$$Logit = g(x) = \log(p(x)(1 - p(x))^{-1}) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k, \quad (12)$$

where $g(x)$ is a linear combination of the independent variables x_1, x_2, \dots, x_k .

The pre-condition for figuring out $P(x)$ is determining parameters α and β_1, \dots, β_k in advance. Using the maximum likelihood method, estimation in logistic regression parameters is chosen [7]. Then, the probability of failure for each input vector x can be calculated according to Eq. (11).

3. Experiment result

The methodology has been applied to evaluate the health condition of a hydraulic pump dynamically. In order to classify the possible root cause, the fault mode analysis was also performed.

3.1. Data acquisition system description

In the experiment, two commonly occurring faults in the hydraulic pump were set, namely, slipper loosening and valve plate wear. The vibration signal of the hydraulic pump was acquired from the end face with a stabilized 528 rpm motor speed and 1000 Hz sampling rate.

3.2. Feature extraction and reduction

First, each vibration signal was processed by local mean decomposition to acquire the product functions. Fig. 2(a) shows the result of the local mean decomposition of the normal condition signal while Fig. 2(b) shows the result of the local mean decomposition of the signal under slipper loosening condition. Second, the skewness coefficients, the kurtosis coefficients and the normalized energy values of the product functions was calculated to form the feature vector. Then PCA was used to determine a subset of feature components. In this case, a 3-dimension feature vector was finally used for the health assessment and fault mode classification.

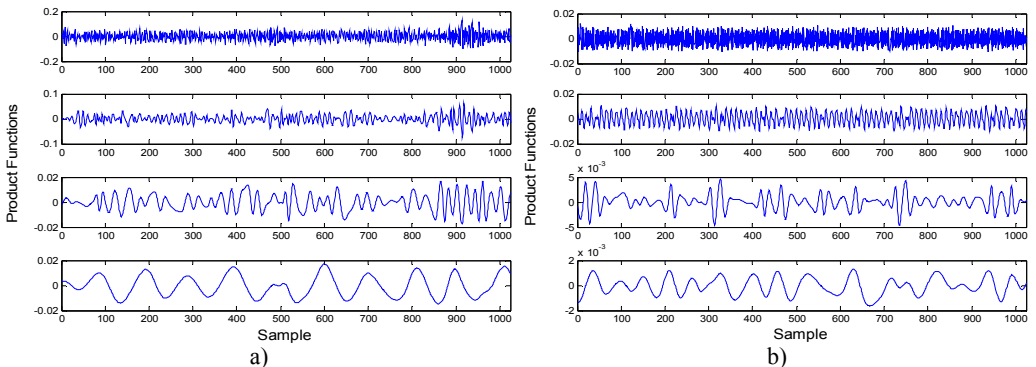


Fig. 2. Local mean decomposition results of normal and slipper loosening condition

3.3. Logistic regression models training

3.3.1. Logistic regression model trained for health assessment

The training data contains 240 sets of data, including 80 sets of normal condition ($P(x) = 0$) data together with 160 sets of fault data ($P(x) = 1$). The parameters α and β_1, \dots, β_3 estimated by the maximum likelihood method was eventually used to obtain the model for performance assessment as LR1.

3.3.2. Logistic regression models trained for root cause classification

Fault mode 1 (valve plate wear): 40 sets of valve plate wear data ($P(x) = 1$) versus 80 sets of normal data ($P(x) = 0$) were used to train the classification model (LR2) for fault mode 1 using the maximum likelihood method;

Fault mode 2 (slipper loosening): 40 sets of slipper loosening data ($P(x) = 1$) versus 80 sets of normal data ($P(x) = 0$) were used to train the classification model (LR3) for fault mode 2 using the maximum likelihood method.

3.4. Validation

For each fault mode, one set data of the hydraulic pump were used for validation.

Set 1 (fault mode 1): 40 sets of data under normal condition versus 80 sets of valve plate wearing data;

Set 2 (fault mode 2): 40 sets of data under normal condition versus 80 sets of slipper loosening data.

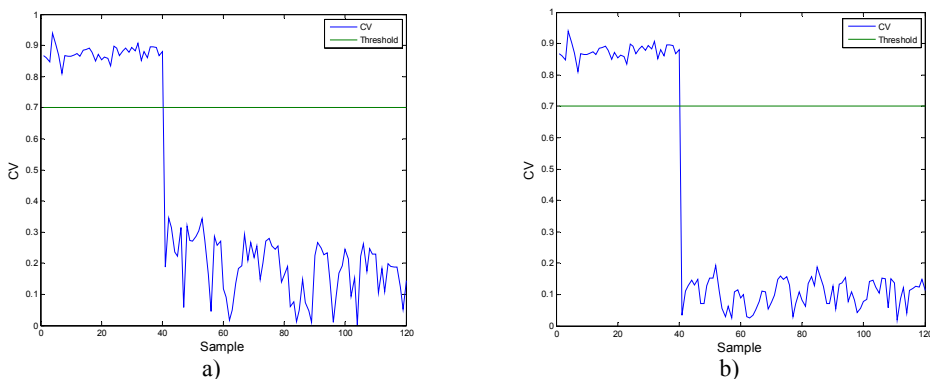


Fig. 3. Health assessment result of two fault modes

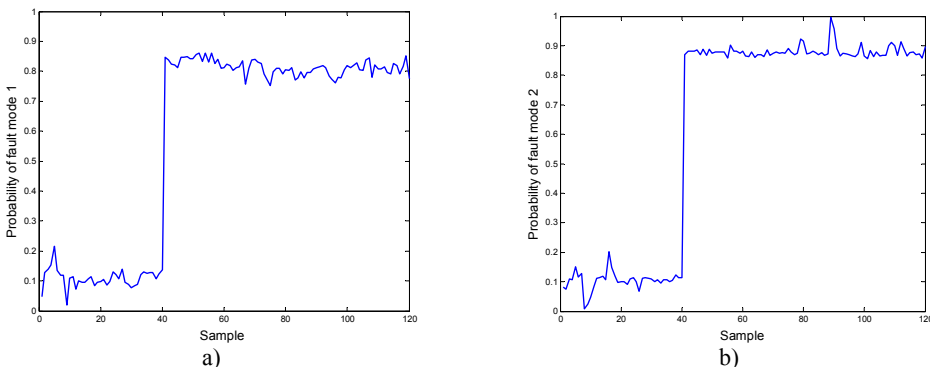


Fig. 4. Probability of fault modes 1 and 2

The confidence value (CV) is defined as $CV = 1 - P(x)$ where $P(x)$ is the probability of failure. When the pump is running normally, CV is very close to 1. However when the pump is going to fail, the CV will approach 0 correspondingly. In the course of the CV falling, as long as the value is less than 0.7, a prefixed threshold, the fault mode classification model will be triggered to input the features into fault classification modes to get the probability of each fault. Fig. 3(a) and (b) shows the result of overall health assessment of the two sets vibration data of hydraulic pump using model LR1. Fig. 4(a) and (b) shows the different fault modes' probability got from LR2 and LR3.

In Fig. 3, both faults can be detected from the CV drops, however, clarifying the difference and the reasons of the drops is still very hard. To solve this problem, a root cause classification module is used. As long as the confidence value is below 0.7, a prefixed threshold, the probability of fault modes will be calculated by inputting the corresponding features into the trained models (LR2 and LR3). In Fig. 4(a) and (b), the probability of fault 1 (valve plate fault) and fault 2 (slipper loosing) is very high from sample 40. Consequently the minor probability points of corresponding fault are showed in Fig. 4.

4. Conclusion

Health assessment and root cause classification of hydraulic pump based on logistic regression has been performed in this paper. LMD combined with PCA can obtain appropriate features from non-stationary signals. LR combined with maximum likelihood technique shows its potential as an effective and efficient tool for health assessment and root cause classification dynamically. The method can not only process non-stationary signals but also stationary signals which makes it generalized and can be applied to other vibration signals of pumps. However, more fault modes should be taken into consideration for better performance on hydraulic pump health assessment and root cause classification.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant No. 61074083, 50705005 and 51105019), as well as the Technology Foundation Program of National Defense (Grant No. Z132013B002).

References

- [1] **Frank P. M.** Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy – a survey and some new results. *Automatica*, Vol. 26, 1990, p. 459-474.
- [2] **Gertler J. J.** Survey of model-based failure detection and isolation in complex plants. *IEEE Control Systems Magazine*, Vol. 8, 1988, p. 3-11.
- [3] **Wang Shaoping, Zhongkui Yuan, Guangqin Yang** Study on fault diagnosis of data fusion in hydraulic pump. *Chinese Journal of Mechanical Engineering*, Vol. 16, Issue 4, 2005, p. 327-331.
- [4] **Smith J. S.** The local mean decomposition and its application to EEG perception data. *Journal of the Royal Society Interface*, Vol. 2, Issue 5, p. 443-454.
- [5] **Cheng J. S., Shi M. L., Yang Y.** Roller bearing fault diagnosis method based on LMD and neural network. *Journal of Vibration and Shock*, Vol. 29, Issue 8, p. 141-144.
- [6] **Djurđjanovic D., Ni J., Lee J.** Time-frequency based sensor fusion in the assessment and monitoring of machine performance degradation. *International Mechanical Engineering Congress and Exposition*, 2002.
- [7] **Yan J., Lee J. J.** Degradation assessment and fault modes classification using logistic regression. *ASME Journal of Manufacturing Science and Engineering*, Vol. 127, 2005, p. 912-914.