A data driven health assessment model for high pressure output pumps in LNG terminals

Yuyun Zeng\textsuperscript{1}, Xiaoshang Wang\textsuperscript{2}, Guangyao Xie\textsuperscript{1} and Jingquan Liu\textsuperscript{1*}

\textsuperscript{1}Department of Engineering Physics, Tsinghua University, Beijing, 100084, China
\textsuperscript{2}Qingdao LNG Co. Ltd., Sinopec, Qingdao, Shandong Province, 266400, China

\textsuperscript{*}Corresponding author’s e-mail: jingquan@tsinghua.edu.cn

Abstract. High pressure output pumps are critical equipment in the vaporization and output system of LNG terminals. Health management helps improve efficiency and reduce cost of the maintenance of high pressure output pumps, thus guaranteeing the efficient productivity of LNG terminals. In order to develop health management system for high pressure output pumps, a data driven health assessment model based on online condition monitoring data of the pumps is proposed. Time domain and frequency domain features are extracted from the monitored vibration signal by statistical analysis and wavelet packet decomposition respectively, and a health index is constructed based on T2 and SPE statistics given by PCA results of the extracted features. The proposed model is validated based on monitoring data of high pressure output pumps in Qingdao LNG terminal. Results show that the calculated health indices are good indicators of the health status of the pumps, and are able to detect potential fault in their early development stages.

1. Introduction

Natural gas industry has been developing rapidly recently in China due to the growing demand on clean and efficient energy and pressure from environmental protection. As important fundamental infrastructures of natural gas industry, liquefied natural gas (LNG) terminals are also developing in high speed accordingly. Totally 21 LNG terminals will be in service by 2020, providing an import capability of nearly 70 million tons of LNG per year.

High pressure output pumps are critical equipment in the vaporization and output system of LNG terminals. Health status of the pumps has a significant influence on production capability of the terminals. According to the operation experience of Qingdao LNG terminal, maintenance of high pressure output pumps suffers from problems of high maintenance cost, high maintenance difficulty and long spare parts procurement cycle. Condition monitoring and health assessment presents a way to deal with these problems via the following two aspects: (1) potential faults can be detected in their early development stages, which helps prevent malfunctions with serious consequences from happening; (2) it can provide information basis for condition based or predictive maintenance decision-making, which results in a more efficient and less costing maintenance and spare parts management process.

There are usually 3 categories of methods for health assessment of rotating machinery: model based[1][2], data driven[3][4], and signal processing based[5][6]. Data driven methods, which build assessment model based on information in historical monitoring data using data mining technology, has become the research focus in health management area due to advantages in flexibility and self-
learning ability, as well as the development of data science in recent years. A data driven model based on principal component analysis method for the health assessment of high pressure output pumps in Qingdao LNG terminal is proposed.

2. Methodology

2.1. Data view and preprocessing

Condition monitoring data of high pressure output pumps in Qingdao LNG terminal fall into 3 categories: (1) process monitoring data stored in the distributed control system (DCS), e.g. inlet and outlet temperature, pressure, vibration amplitude, etc.; (2) electric parameters of motors stored in the power management system; (3) vibration monitoring data stored in the vibration monitoring system. The vibration monitoring data are collected by two accelerometers mounted on the drive end bearing for each pump on X and Y directions. Every 10 minutes, a vibration waveform containing 2048 data points with sampling frequency set as 10 kHz is acquired by each accelerometer.

Operation experience show that the main cause that leads to breakdown of high pressure output pumps in Qingdao LNG terminal is wear of bearings, which are usually reflected in vibration monitoring among all condition monitoring measurements. Thus, vibration data are utilized to build the health assessment model for high pressure output pumps.

Firstly, features that represent the development of faults are extracted from the original monitored vibration data. Figure 1 shows spectrums of monitored vibration signals on 3 time points during the development of a bearing wear fault that lead to a pump breakdown in April, 2018. As can be seen, vibration energy concentrates around 2 kHz in normal operation condition of the pump, while energy of frequency bands over 3 kHz rise with the development of fault. Vibration energy of different frequency bands can therefore be used as features for the health assessment model.

Wavelet packet decomposition (WPD) is used to extract energy of frequency bands from monitored vibration waveforms. A $M$-layer WPD decomposes the signal into $2^M$ equal width frequency bands by applying a series of high pass and low pass filters[7]. WPD coefficients are calculated recursively by:
where $d$ represents WPD coefficients, $j$ and $n$ represent node number, $l$ and $k$ represent layer number, and $g$ and $h$ represent coefficients of the high pass and low pass filters, respectively.

Features are then constructed as energy ratios of the frequency bands which are calculated based on the sum of square of the corresponding WPD coefficients and a normalization process[8]. Additionally, time domain statistics that show significant correlation with the development of fault including root mean square, kurtosis, and margin factor also extracted for the health assessment model.

2.2. Model construction based on principal component analysis

Principal component analysis (PCA) based health assessment model represents health status of the pumps by points in the transformed feature space. Healthy status is defined by an area in the feature space based on historical data, and the health index is calculated by comparing features extracted from the current monitoring data to the defined healthy area.

Specifically, the health assessment model is constructed by the following two stages:

(1) Offline stage: definition of healthy status.

Select a segment of data in historical monitoring database where the pump works in healthy status (e.g. after maintenance or replacement of spare parts), and extract the features described in section 2.1 to form the feature matrix $X$. Then perform PCA on $X$ (refer to [9] for the specific process of PCA). Reserve $k$ components so that the explained variance ratio is no less than 90%, and form the truncated eigenvalue vector $\Lambda = [\lambda_1, \lambda_2, ..., \lambda_k]^T$ and transformation matrix $P = [p_1, p_2, ..., p_k]^T$, where $\lambda_i$ and $p_i$ are the $i$th maximum eigenvalue and the corresponding eigenvector of the covariance matrix of $X$, respectively.

Transform $X$ into the principal component space by the transformation matrix:

$$Y = PX$$

where $Y$ represents the transformed data in the principal component space, and the healthy status can be defined by the statistic distribution of data in $Y$.

(2) Online stage: health index calculation.

Once online monitoring data are acquired and the corresponding features are extracted, health index are to be calculated in real time based on the comparison between online features and the healthy status defined above. The comparison is conducted based on two statistics, i.e. T2 and SPE statistics.

T2 statistics reflect the degree that the current monitoring data deviates from the defined health status, and is calculated as:

$$T2 = x^T P^T S^{-1} P x$$

where $x$ represents the feature vector of online monitoring data, and $S$ represents a diagonal matrix whose diagonal elements are the truncated eigenvalue vector $A$, i.e. $S = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_k)$.

SPE statistics represents the error of PCA transformation (2) on the online monitoring data:

$$SPE = x^T (I - P^T P)x$$

where $I$ represents the identity matrix.

Finally, in order to get a health index within range [0, 1], a sigmoid mapping is applied on the calculated T2 and SPE statistics:

$$HI = \left(1 + e^{w_x^T x + b} \right)^{-1}$$
where HI represents the health index, \( z \) represents the vector of T2 and SPE statistics, i.e. \( z = [T2, SPE]^T \), and \( w \) and \( b \) are model parameters. Model parameters are set based on the statistics limits derived by hypothesis test so that the health indices calculated would be around 1 for healthy status and around 0 for faulty status.

3. Result and discussion

Monitoring data during the development of 2 excessive vibration related pump breakdowns were collected from the historical monitoring database in Qingdao LNG terminal for model validation. Figure 2 shows the evolution of vibration amplitude in the test cases.

![Figure 2: Vibration amplitude in two test cases: (a) case 1; (b) case 2.](image)

Time domain and energy ratio features were firstly extracted from the original vibration signals. For energy ratio features, a 3-layer WPD using dmey wavelet was applied, decomposing the vibration signals into 8 frequency bands. Energy ratio of all 8 frequency bands were calculated as features for the health assessment model. Figure 3 shows an extracted energy ratio feature (energy ratio of the 3rd frequency band) during the development of two test cases, respectively. Significant drop were spotted around the time point that fault occurred, showing that the extracted features are good indicators of the development of fault in the pumps.

![Figure 3: Energy ratio of the 3rd frequency band in test cases: (a) test case 1; (b) test case 2.](image)

With data before the occurrence of fault in both cases (i.e. before September 2017 in case 1 and before November 2017 in case 2) selected as reference data to define the healthy status, the PCA based health assessment model was constructed for each pump following the steps described in section 2.2. Calculated T2 and SPE statistics for both cases are shown in figure 4. As can be seen, T2 and SPE statistics fluctuated in a small range when the pump works in normal operation conditions. After fault occurred, however, both statistics increased dramatically until breakdown of the pumps triggered by excessive vibration. It is thus proved that T2 and SPE statistics given by the PCA based health assessment model reflect health status of the pumps well.
Figure 4. Calculated T2 and SPE statistics for both test cases: (a) case 1, T2; (b) case 1, SPE; (c) case 2, T2; (d) case 2, SPE.

Health indices of high pressure output pumps were than calculated by sigmoid mapping of T2 and SPE statistics. The calculated health indices for both test cases are shown in figure 5. With the development of faults, health index decreases from around 1 to around 0, indicating that the pumps moves from healthy status to faulty status. Dramatic decrease on health index is spotted on around December 13th, 2017 for test case 1, implying that fault has been detected in the corresponding pump and need to be paid special attention to. However, as shown in figure 2(a), at the same time vibration amplitude (which is displayed in DCS) has not experience significant increase and is still far from the breakdown threshold. The same tendency can be spotted on around February 1st, 2018 for test case 2. It is therefore proven that the calculated health index is sensitive to anomalies in the health status of corresponding pumps, and can detect potential faults in their early development stages.

Figure 5. Calculated health indices in two test cases: (a) case 1; (b) case 2.

4. Conclusion

In order to improve maintenance efficiency and reduce maintenance cost for high pressure output pumps in LNG terminals, a data driven health assessment model based on online condition monitoring data is proposed. Time domain and frequency domain features are firstly extracted from the original
vibration signals via statistic analysis and WPD, respectively. PCA is utilized on the extracted features to define a healthy status and calculate T2 and SPE statistics that reflect the deviation from the defined healthy status, and health index is then constructed by a sigmoid mapping of T2 and SPE statistics. The proposed model is validated based on actual monitoring data of high pressure output pumps in Qingdao LNG terminal. Results show that the calculated health indices can represent the health status of the pumps, and are able to detect potential faults in their early development stages, which can be valuable in maintenance management of the pumps.

Acknowledgments
This study is sponsored by the project “Research on health assessment and maintenance strategy for critical equipment in LNG terminals” of Qingdao LNG Co. Ltd., Sinopec.

References
[1] Feng, Z., Qin, S., Liang, M. (2016) Time-frequency analysis based on Vold-Kalman filter and higher order energy separation for fault diagnosis of wind turbine planetary gearbox under nonstationary conditions. Renewable Energy, 85: 45-56.
[2] Ma, D., Ren, S., Liu, K., et al. (2018) Fault diagnosis of rolling bearing based on Bayesian network and fuzzy evaluation. Journal of Harbin University of Science and Technology, 23: 114-118.
[3] Liu, J., Seraoui, R., Vitelli, V., et al. (2013) Nuclear power plant components condition monitoring by probabilistic support vector machine. Annals of Nuclear Energy, 56: 23-33.
[4] Qi, G., Zhu, Z., Erqinhu, K., et al. (2018) Fault-diagnosis for reciprocating compressors using big data and machine learning. Simulation Modelling Practice and Theory, 80: 104-127.
[5] Porocnik, P., Govekar, E. (2017) Semi-supervised vibration-based classification and condition monitoring of compressors, 93: 51-65.
[6] Wang, J., Bu, Q., Zhao, H., et al. (2019) Fault diagnosis of reciprocating compressor on the resonance-based sparse signal decomposition with optimal Q-factor. Journal of Mechanical Strength, 41: 557-561.
[7] Zhao, P., Gu, L., Sun, Y., et al. (2019) Pressure pulsation extraction and analysis based on wavelet packet. Machinery Design & Manufacture, 7: 161-164.
[8] Wang, H., Zhou, J., Fang, Q. (2019) Feature extraction method of loudspeaker abnormal sound based on wavelet packet decomposition and sample entropy. Journal of Xi’an Polytechnic University, 33:57-62.
[9] Zhao, Y., Wu, M., He, L. (2019) Abnormal diagnosis of machine tool health status based on PCA statistics. Mechanical Engineering & Automation, 5: 164-166.