Machine learning based health assessment model for high pressure output pumps in LNG terminals

Yuyun Zeng¹, Wenbo Zhu², Jingjie Yang², Guangyao Xie¹ and Jingquan Liu*¹

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

*Corresponding author’s e-mail: jingquan@tsinghua.edu.cn

Abstract. High pressure output pumps are one of the critical equipment in LNG terminals. Since the health condition of high pressure output pumps has a direct influence on production capability of the terminal, health assessment for these pumps in real time plays an important role on guaranteeing efficient productivity of LNG terminals. Using condition monitoring data, a machine learning based health assessment model for high pressure output pumps is proposed. Health features are constructed based on time domain statistical analysis and wavelet packet decomposition, and a SVR model is trained to calculate a health index from extracted features. Actual operating data in Qingdao LNG terminal are used for model validation. Results show that the calculated health indices are sensitive to faults and anomalies of the pumps, and are good indicators of pump health status. The proposed model also shows capability of early warning for some sudden failures, which can be valuable in the operation and maintenance management of LNG terminal equipment.

1. Introduction

Driven by the growing demand on clear and efficient energy, liquefied natural gas (LNG) industry has recently become the most rapidly growing area in energy chemical engineering industry in China. LNG terminals are fundamental infrastructures in LNG industry. High pressure output pumps are critical equipment in the vaporization and output system of LNG terminals. Their operating reliability has a direct influence on the production capability of the terminals. Qingdao LNG terminal has 6 high pressure output pumps in service. Operation experience shows that 9 corrective maintenance actions have been conducted in the last 4 years, costing nearly 4 million yuan. Improving the operation and maintenance management of critical equipment has thus become an urgent demand in LNG terminals.

Condition monitoring and health management is state-of-art in equipment maintenance management area. Through health assessment based on condition monitoring data, anomalies in equipment operation can be detected effectively, and the health status of equipment can be acquired in real time, which helps prevent significant malfunctions from happening, as well as providing information basis for condition based or predictive maintenance decision-making, resulting in a more efficient and less costing equipment maintenance process. For rotating machinery including pumps, health assessment methods mainly fall into 3 categories: model based[1][2], data driven[3][4] and signal processing based[5][6]. With the rapid development of data science in recent years, data driven methods has become the research focus in health management area due to their advantages in flexibility and self-learning ability. Based on data driven framework, a machine learning based online health assessment model is proposed for high pressure output pumps in Qingdao LNG terminal.
2. Methodology

2.1 Condition monitoring data on site
Four high pressure output pumps in the phase-I project of Qingdao LNG terminal are taken as research object. Condition monitoring data are distributed in 3 different information systems on site:
- Distributed control system (DCS): contains most process monitoring data, e.g. inlet and outlet temperature, pressure, liquid level, etc.
- Power management system: contains electric parameters of motors.
- Vibration monitoring system: two accelerometers are installed on the drive end bearing for each pump, monitoring vibration on X and Y direction. Vibration signal is measured every 10 minutes. Each sample contains 2048 data points with sampling frequency set as 10 kHz.

According to maintenance records, the main cause of high pressure output pump breakdown was wear of bearings due to defects on axial balance. Operation experience and data exploring both show that wear of bearings is usually reflected in vibration monitoring, thus the proposed health assessment model is constructed mainly based on vibration monitoring data of the pumps.

2.2 Feature extraction
Features are input variables of the health assessment model, and have significant influence on the performance of the constructed model. In order to achieve better model performance, features that can reflect the development of the corresponding failure mode should be extracted from original monitoring data through the combination of mechanism analysis, experience judgement and data analysis.

![Monitored spectrum on 3 time points before breakdown](image)

Figure 1. Monitored spectrum on 3 time points before breakdown: (a) 2018-01-05 00:10; (b) 2018-03-01 00:10; (c) 2018-03-31 00:10.

Figure 1 shows monitored spectrums of bearing vibration on 3 time points before the breakdown of a pump in April, 2018. It can be seen that vibration energy is mainly concentrated in the frequency band around 2 kHz under normal operation condition. With the development of wearing, energy of frequency bands over 3 kHz rise gradually, until pump breakdown is triggered when vibration amplitude
reaches the setting value. Vibration energy in different frequency bands can thus be the features that indicate the development of concerned fault of high pressure output pumps.

Wavelet packet decomposition (WPD) is utilized to extract frequency band energy from monitored vibration signals. WPD is time-frequency domain signal processing method developed from wavelet analysis[7]. By applying a series of high pass and low pass filters, signals are decomposed into $2^N$ frequency bands (N is the number of layers of WPD). The recursive formula for coefficients in WPD can be written as[8]:

\[
\begin{align*}
    d_{k}^{l+1,2n} &= \sum_{j} h_{2l-k} d_{j}^{l,n} \\
    d_{k}^{l+1,2n+1} &= \sum_{j} g_{2l-k} d_{j}^{l,n}
\end{align*}
\]

(1)

where $d$ represents WPD coefficients, $j$ and $n$ represent node number, and $l$ and $k$ represent layer number.

The square of WPD coefficients has a dimension of energy[9], so the sum of square of coefficients represent a measure of energy for the corresponding frequency band. The WPD band energy ratio features are further constructed by a normalization process, i.e., divided by the total vibration energy. Along with WPD band energy ratios, time domain statistics that show significant correlation with the fault development process including root mean square, kurtosis, margin factor etc. are also extracted as features for the health assessment model.

2.3 Machine learning based health assessment model

Health assessment model takes extracted features as input, and map them to a health index that directly reflect the health status of the corresponding pump. Health indices usually range from 0 to 1. A value around 0 indicates normal operation status, while a value near 1 corresponds to significant fault status.

Support vector regression (SVR) is used to construct the mapping from features to the health index. SVR can be seen as a kind of nonlinear kernel regression model which has the following form[10]:

\[
y = \sum_{i} (\alpha_{i} - \alpha_{i}^{*}) k(x_i, x) + b
\]

(2)

where $x$ represents the input vector, i.e. extracted features, $y$ represents the output value, i.e. the health index, $x_i$ ($i = 1, 2, \ldots, M$) represents the training data points, $\alpha_i$, $\alpha_i^{*}$ and $b$ are model parameters that are to be determined by the model training process, and $k(x_i, x)$ is the kernel function. Radial basis function (RBF) is used as kernel function in the health assessment model:

\[
k(x_i, x) = \exp \left( -\frac{\|x_i - x\|^2}{\gamma} \right)
\]

(3)

where $\gamma$ is a model parameter to be determined by the training process.

The SVR based health assessment model is constructed by the following steps:

- Data collection: collect historical monitoring data which cover both normal operation status and status near breakdown. Extract WPD energy ratio and time domain features from the collected data.
- Labelling: label data in normal operation status as 0, and data near breakdown as 1. Use the extract features as input and the corresponding label as output to build the training set.
- Model training: use SVR training algorithm to determine model parameters in equations (2) and (3) based on the training set.

3. Result and discussion

Totally 6 excessive vibration related breakdown have happened in historical monitoring data of the 4 high pressure output pump in Qingdao LNG terminal. 4 of them were used to build the training set for
As shown in figure 2, the test cases represent two typical situations of pump breakdown by excessive vibration: in test case 1, vibration amplitude increased gradually over a period of time, while vibration amplitude just went up suddenly to over breakdown threshold in case 2.

Time domain analysis and WPD were used for the feature extraction process. A 3-layer WPD was applied to calculate energy ratio features. The monitored vibration signals were decomposed into 8 successive frequency bands, and energy ratio of each frequency band was calculated as input features for the health assessment model. Figure 3 and figure 4 show the extracted energy ratio of the 2nd and 5th frequency band for test case 1 and case 2 respectively. As can be seen in figure 3, with the development of fault in test case 1, energy ratio of the 2nd band dropped significantly and energy ratio of the 5th band increased correspondingly, proving that the extracted features are good indicators of fault development in the pumps.

As shown in figure 4, before the pump breakdown caused by sudden excessive vibration happened around July 19, 2017, drop in energy ratio of the 2nd band and increase in energy ratio of the 5th band were spotted from July 8 to July 12, which implies that the extracted features can be symptoms of the upcoming failure.

Using extracted energy ratio and time domain features as input variables, a SVR health assessment model was trained using the process mentioned in section 2.3. Health indices in test cases were calculated using the trained model, and the results are shown in figure 5. For case 1 (i.e., figure 5(a)), the health index rose dramatically from near 0 to near 1 on around January 27, 2018, and kept in high value range afterwards, implying that fault has occurred on the corresponding pump and should be paid special attention to. At the same time, the vibration amplitude shown in figure 2(a) (which is displayed in DCS) had not experienced significant increase and was still far from breakdown threshold. Thus it is proven that the proposed health assessment model has high sensitivity towards anomalies in operation condition of the pumps, and is able to detect faults in their early development stages.
For the sudden failure case (i.e. test case 2) shown in figure 5(b), the calculated health index stayed around 0 in normal operation status. However, large variation in the health index was spotted on a few days before failure happened with the largest value reached around 0.7, which can be seen as an early warning for the upcoming failure.

4. Conclusion
A machine learning based health assessment model is proposed in order to improve the maintenance management process of high pressure output pumps in LNG terminals. The proposed model analyzes related on-site condition monitoring data, and assess the health status of corresponding pumps in real time. Time domain analysis and WPD are used to extract time and frequency domain features from the original monitored vibration signals, and a SVR model is trained to map the extracted features to a health index that directly reflects equipment health status. Actual operating data of high pressure output pumps in Qingdao LNG terminal are collected for model validation. Results show that the calculated health indices are good indicators of pump health status, and are sensitive to faults and anomalies. The proposed model can detect potential faults in their early development stages, and also shows capability of early warning for some sudden failures. The proposed model can also be extended to other equipment in LNG terminals.

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] Widodo, A., Yang, B. (2007) Support vector machine in machine condition monitoring and fault diagnosis. Mechanical Systems and Signal Processing, 21: 2560-2574.

[4] Zhou, Y. (2018) Research on Performance Degradation Assessment and Remaining Useful Life Prediction Methods of Rolling Bearings. South China University of Technology, Gangzhou.

[5] Shi, L., Zhang, Y., Mi W. (2011) Application of Wigner-Ville-distribution-based spectral kurtosis algorithm to fault diagnosis of rolling bearing. Journal of Vibration, Measurement & Diagnosis, 31: 27-31.

[6] Wang, X., Jiang, Q. (2014) Application study of wavelet analysis in steam turbine vibration fault diagnosis. Journal of Electirc Power, 29: 80-83.

[7] Zhao, P., Gu, L., Sun, Y., et al. (2019) Pressure pulsation extraction and anlysis 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] He, Q., Yao, C., Chen, G., et al. (2012) Feature extraction method of hydraulic pump vibration signal based on singular value decomposition and wavelet packet analysis. Journal of Data Acquisition & Processing, 27: 241-247.

[10] Smola, A.J., Schölkopf, B. (2004) A tutorial on support vector regression. Statistics and Computing, 14: 199-222.