



**HAL**  
open science

# A Study on the Diagnostics Method for Plant Equipment Failure

Minyoung Seo, Hong-Bae Jun

► **To cite this version:**

Minyoung Seo, Hong-Bae Jun. A Study on the Diagnostics Method for Plant Equipment Failure. IFIP International Conference on Advances in Production Management Systems (APMS), Sep 2019, Austin, TX, United States. pp.701-707, 10.1007/978-3-030-30000-5\_85 . hal-02419258

**HAL Id: hal-02419258**

**<https://inria.hal.science/hal-02419258v1>**

Submitted on 19 Dec 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# A study on the diagnostics method for plant equipment failure

Minyoung Seo<sup>1</sup> and Hong-Bae Jun<sup>2</sup>[0000-0001-5856-7083]

Puzzle systems co., Data Business Unit, Seoul 04066, South Korea  
nepoentrood@gmail.com

Hongik University, Dept. of Industrial Engineering, Seoul 04066, South Korea  
hongbae@gmail.com

**Abstract.** Recently, in the era of the Fourth Industrial Revolution, the rapid development of ICT(Information and Communication Technology) and IoT(Internet of Things) technology have been actively applied to collect and utilize the status data of plant equipment during their operation period. With these technologies it is very important to keep the availability and reliability of the equipment during its usage period without any interruption or failure. In this vein, the CBM(Condition Based Maintenance) or PHM (Prognostics and Health Management) policy which carries out maintenance activities based on the condition of the equipment has been increasingly applied to the plant industry. Although it has a high potential to derive the important value from operation data of plant equipment through data analytics, research on data analytics in the plant industry is still known as an early stage. In this study, we briefly introduce a method to diagnose the fault state of the equipment by detecting patterns related to the failure modes of equipment based on gathered sensor data. To develop the method, we apply the well-known clustering/classification algorithms and text mining and information retrieval method. In a case study, we apply the proposed method and show its possibility throughout preliminary experiments.

**Keywords:** Failure diagnosis · Plant equipment · Maintenance · CBM.

## 1 Introduction

Generally, a plant is the complex system consisting of a number of highly reliable equipment. It has been in operation for decades, usually through years of design and construction. Recently, the rapid development of ICT(Information and Communication Technology) and IoT(Internet of Things) technology have been actively applied to collect and utilize status data of major plant equipment during their lifecycle. With these technologies it is very important to keep the availability and reliability of the equipment during its usage period without any interruption or failure. During the plant operation, any small accident could lead to a lot of human and material damage at the industrial and power plants such as nuclear power plants, petroleum refineries, chemical plants, etc.. For example,

in the case of an energy plant handling hazardous chemicals with high risk, a leak accident can cause catastrophic damage to workers in industrial facilities as well as to civilians outside the industrial facility. Therefore, it is necessary to develop the diagnostics method for plant equipment failure in order to prevent such an accident in advance.

In this vein, the CBM(Condition Based Maintenance) or PHM (Prognostics and Health Management) policy which carries out maintenance activities based on the condition of the equipment has been increasingly applied to the plant industry. Although it has a high potential to derive the important value from operation data of plant equipment during its operation period through data analytics, research on data analytics in plant industry is still known as an early stage.

In this study, we briefly introduce a method to diagnose what kinds of faults will appear in the future by detecting fault patterns related to plant equipment based on gathered sensor data under the CBM policy. We apply the TF-IDF(Term Frequency-Inverse Document Frequency) technique frequently used in text mining and information retrieval to detect anomalous patterns related to equipment failure types. The TF-IDF value is a value indicating how important a word is in a specific document, and many search engines use it for evaluating the importance of words existing in a document. In this study, we used the TF-IDF technique to extract patterns of faults that were related to the fault type among the patterns that occurred before the specific fault. In addition, we applied the SVM (Support Vector Machine) to establish the fault diagnosis model. To evaluate the proposed method, we have carried out a case study for the centrifugal pump and found out the possibility of the proposed method throughout preliminary experiments.

This study is organized as follows: Section 2 reviews the relevant previous works. Section 3 briefly introduces the overview of the proposed method. Section 4 summaries the preliminary test result of the case study. Finally, section 5 concludes the paper with a short summary.

## 2 Previous works

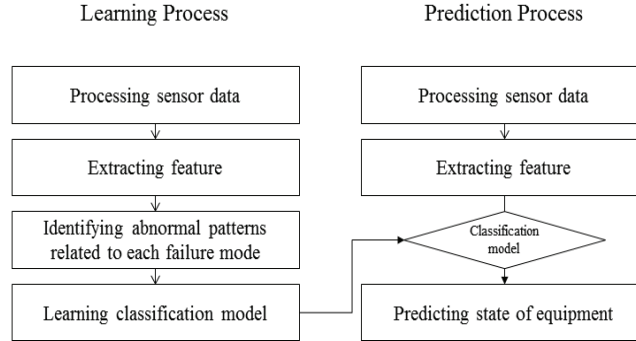
In general, maintenance activities are divided into ‘breakdown maintenance’ and ‘preventive maintenance’. Under the breakdown maintenance policy, maintenance activities are performed after recognizing the failure. On the other hand, the preventive maintenance performs maintenance activities based on the time or equipment condition before the failure occurs. The preventive maintenance is again classified into TBM(Time Based Maintenance) and CBM(Condition Based Maintenance). The TBM is a method of regularly checking and preserving the target equipment with a fixed period of time. The CBM is to carry out maintenance activities based on the condition of the equipment. This study focuses on the CBM policy. In the CBM, one main challenging issue is the failure diagnostics.

So far there have been several research works related to CBM studies and fault diagnosis methods. For example, Bunks *et al.* (2000)[1] proposed a state-based conservation method using the HMM(Hidden Markov Model) and conducted case studies using data collected from the Westland helicopter gearbox. Djurdjanovic *et al.* (2003)[3] presented the concept of a watchdog agent to perform CBM on mechanical equipment using sensors and wireless internet technologies that collect various data. Dong *et al.* (2007)[4] diagnosed the state of the equipment based on the data collected from various sensors with the HSMM(Hidden Semi-Markov Model) technique, which is different from the HMM with fixed state transition probability, that changes the state transition probability with time. Wang *et al.* (2007)[9] used a partly-linearized neural network to extract features using a continuous wavelet transform technique for centrifugal pumps for diagnosing equipment failure. Furthermore, Li *et al.* (2009)[6] proposed a method for predicting the deterioration of a gas turbine engine using a combination of linear regression and quadratic regression. Fei *et al.* (2009)[5] proposed a methodology to diagnose the failure of power transformation based on the information of gas collected from equipment using SVMG (Support Vector Machine with Genetic Algorithm) method which determines parameters of SVM using genetic algorithm did. Zhao *et al.* (2013)[10] proposed a generic intelligent fault detection and diagnosis strategy that simulates the actual diagnostic thinking of chiller experts. The diagnosis is performed by a three-layer DBN. Muralidharan *et al.* (2014)[8] studied the method of extracting features from centrifugal pump by using continuous wavelet transform technique from vibration data and diagnosing equipment failure by SVM. In addition, Moon and Kim (2014)[7] proposed a system that could efficiently diagnose mechanical failures such as mass imbalance and axis alignment failure frequently occurred in wind power generation system by using artificial neural network and wavelet transform. Cai *et al.* (2014)[2] applied an expanded BN model to the failure prediction process of a complex system under uncertainty, and built the corresponding model from a data mining algorithm based on the divide-and-conquer principle.

In this study, we first propose a method to diagnose the state of equipment using TF-IDF technique, which is mainly used in information retrieval and text mining. In many studies, wavelet transform was used to extract anomalous patterns from univariate (mainly vibration) data. However, in this study, features are defined for multivariate (vibration, temperature, etc.) data. The abnormal pattern is extracted using the TF-IDF technique and clustering methods. These points are different from the previous studies.

### 3 Approach

In this section, we briefly introduce the approach to diagnose the failure type of the equipment based on sensor data collected from plant equipment. Figure 1 shows the overview of the method proposed in this study.



**Fig. 1.** Overview of the proposed approach

As shown in Figure 1, the proposed method is largely divided into two parts. One is the learning process to diagnose the condition of the equipment and the other is the process of predicting the state of the equipment using the learned model.

Before assuming a concrete explanation of the method, the assumptions of this study are as follows. First, the state of the equipment consists of normal and fault types. Second, there must be at least two states of equipment. Third, there is at least one pattern in the state of each equipment.

The overall procedure for learning is as follows. The first step is to remove the missing values and perform the normalization process to refine the collected sensor data. In the second step we divide the data interval and extract the data summation value of gathered sensor data in each interval to construct the feature vector. In this study, the feature is defined as a data summation value (e.g., average, standard deviation, maximum value, minimum value) in a specific interval, and a feature vector is defined as a set of data summation values in a specific interval.

In the third step, we find the patterns associated with failure types. The pattern means the clustered feature vectors and the abnormal pattern means the pattern related to a certain failure type. The normal pattern means the other patterns except the abnormal patterns. To find the patterns among feature vectors, the  $k$ -means clustering method is applied. Then, to identify the pattern associated with the failure type, the PF-IFF (Pattern Frequency – Inverse Failure mode Frequency) value based on the TF-IDF technique between the pattern and the equipment failure type is calculated with the following formulae.

$$PF(p, f) = 0.5 + \frac{0.5 \times f(p, f)}{\max\{f(c, f) : c \in f\}} \quad (1)$$

$$IFF(p, f) = \log\left(1 + \frac{|F|}{|(f \in F : p \in f)|}\right) \quad (2)$$

$$PF\_IFF(p, f, F) = PF(p, f) \times IFF(p, F) \quad (3)$$

Here  $PF$  denotes the degree of occurrence of a particular pattern  $p$  in a particular failure mode  $f$ , and  $IFF$  denotes the degree of occurrence of a particular pattern  $p$  in the failure mode set  $F$  in common. Thus the  $PF\_IFF$  value is a score indicating how important a pattern is in a particular failure mode when there are multiple failure modes.

Next, the patterns with a high  $PF\_IFF$  value in a particular failure type are determined as the failure patterns associated with that failure type. In the final step, the relationships between the  $PF\_IFF$  values of the patterns from feature vectors and their matching failure types are studied using the SVM technique. After the learning model is built, the state of the equipment is diagnosed by inputting a feature vector composed of data newly gathered into the classification model learned in the learning process.

## 4 Case study

In this study, a case study was conducted based on the data collected from centrifugal pump of a Korean company to verify the effectiveness of the proposed method. The centrifugal pump is an energy conversion device that transfers mechanical energy to the working fluid. The data to be covered in the case study are the centrifugal pump state data collected from June 2013 to July 2015 at intervals of 10 minutes, and the data attributes include the vibration, temperature, speed, and axial movement of each of 15 pump positions. The types of failures to be considered in this study are ‘leakage’, ‘misalignment’, and ‘bearing damage’. In this study, the data analytics language  $R$  was used to implement the proposed method.

The case study was conducted by dividing the training data and the experimental data at a ratio of 4:1. The classification model that can diagnose the state of the equipment with learning data has been built. Based on it, we evaluated the performance of the classification model by predicting the state of equipment with experimental data.

The performance criterion for the proposed method was set as AUC(Area Under the Curve) in the ROC(Receiver Operating Characteristic) Curve. The closer the AUC is to 1, the better the performance.

**Table 1.** Preliminary test result

Method	Normal	Bearing damage	Leakage	Misalignment
PF\_IFF used	0.957	0.977	0.930	0.765
PF\_IFF not used	0.722	0.593	0.739	0.689

Table 1 compares the AUC values of two kinds of models: One is our approach that uses *PF\_IFF*. The other is the method that does not use *PF\_IFF*. Preliminary experimental results showed that the diagnostic performance of each condition of the equipment is 0.957 for Normal, 0.977 for Bearing Damage, 0.930 for Leakage and 0.765 for Misalignment. Thus, the *PF\_IFF* method proposed in this study showed the better performance in all cases in diagnosing the equipment condition than the unused method.

## 5 Conclusion

In this study, we have briefly introduced a method for diagnosing failure types based on data collected from sensors in plant equipment. The *PF\_IFF* technique using TF-IDF used in keyword extraction of the text mining and the information retrieval field was used to extract anomalous patterns for each type of equipment and SVM was used to diagnose the failure pattern. The TF-IDF value is a value indicating how important a word is in a specific document. The case study was conducted based on the data about vibration, temperature, and axial movement collected from the centrifugal pump. AUC was used as a performance evaluation criterion of the model, and the performance of the diagnosis model of the equipment using the proposed *PF\_IFF* was evaluated. Preliminary test results showed that the diagnostics performance of the equipment with our approach is reasonable.

There are two limitations of this study. First, although the proposed method predicts what kind of failure will occur when an anomaly pattern is found in the equipment, but it cannot estimate when the actual failure will occur. The second is that certain patterns that appear could be diagnosed unexpectedly as normal if they occur in all types of failures in the equipment, despite the fact that there is an abnormality in the condition of the equipment in a particular section. These must be future research issues.

## Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2017R1D1A1B03031633).

## References

1. Bunks, C., McCarthy, D., Al-Ani, T.: Condition-based maintenance of machines using hidden Markov models. *Mechanical Systems and Signal Processing* **14**(4), 597–612 (2000).
2. Cai, Z., Si, W., Si, S., Sun, S.: Modeling of Failure Prediction Bayesian Network with Divide-and-Conquer Principle. *Mathematical Problems in Engineering*, 1–8 (2014).

3. Djurdjanovic, D., Lee, J., Ni, J.: Watchdog Agent—an infotonics-based prognostics approach for product performance degradation assessment and prediction. *Advanced Engineering Informatics* **17**(3), 109–125 (2003).
4. Dong, M., He, D.: Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis. *European Journal of Operational Research* **178**(3), 858–878 (2007).
5. Fei, S.W., Zhang, X.B: Fault diagnosis of power transformer based on support vector machine with genetic algorithm. *Expert Systems with Applications* **36**(8), 11352–11357 (2009).
6. Li, Y.G., Nilkitsaranont, P.: Gas turbine performance prognostic for condition-based maintenance. *Applied energy* **86**(10), 2152–2161 (2009).
7. Moon, D.S., Kim, S.H.: Development of intelligent fault diagnostic system for mechanical element of wind power generator. *Journal of Korean Institute of Intelligent Systems* **24**(1), 78–83 (2014).
8. Muralidharan, V., Sugumaran, V., Indira, V.: Fault diagnosis of monoblock centrifugal pump using SVM. *Engineering Science and Technology, an International Journal* **17**(3), 152–157 (2014).
9. Wang, H.Q., Chen, P.: Fault diagnosis of centrifugal pump using symptom parameters in frequency domain. *The CGIR Journal* **9**, 1–14 (2007).
10. Zhao, Y., Xiao, F., Wang, S.: An intelligent chiller fault detection and diagnosis methodology using Bayesian belief network. *Energy and Buildings* **57**, 278–288 (2013).