Journal of Computer and Communications, 2020, 8, 36-44
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Research on Fault Prediction for Marine Diesel Engines

Zhengyang Qi, Yunsong Qi, Guangpeng Hu

School of Computer, Jiangsu University of Science and Technology, Zhenjiang, China
Email: 296913335@qq.com, mailqys@163.com

How to cite this paper: Qi, Z.Y., Qi, Y.S. and Hu, G.P. (2020) Research on Fault Prediction for Marine Diesel Engines. *Journal of Computer and Communications*, **8**, 36-44.
https://doi.org/10.4236/jcc.2020.88004

Received: July 29, 2020
Accepted: August 23, 2020
Published: August 26, 2020

Abstract

Condition-based maintenance based on fault prediction has been widely concerned by the industry. Most of the contributions on fault prediction are based on various sensor data and mathematical models of the equipment. The complexity of the model and data signal is the key factor affecting the practicability of the model. In addition, even for the same type and batch of equipment, the manufacturing process, operation environment and other factors also affect the model parameters. In this paper, a series event model is conducted to predict the fault of marine diesel engines. Numerical example illustrates that the proposed event model is feasible.

Keywords

Condition-Based Maintenance, Series Events, Fault Prediction, Marine Diesel Engine

1. Introduction

Equipment maintenance activities can be roughly divided into three categories: corrective maintenance, planned maintenance and condition-based maintenance (CBM). In corrective maintenance, the equipment has failed, and the purpose of maintenance activities is to repair the failure to quickly recover the system [1] [2] [3]. This process passively waits for the failure and then fixes it. During scheduled maintenance, check the machine regularly and replace the appropriate parts according to the fixed schedule. Therefore, it is also called time-based maintenance [4]. CBM is a flexible method to determine whether maintenance activities are needed or not according to the actual health status of the equipment. It emphasizes the prediction of the possible faults of the equipment in advance to make maintenance strategies and improves the operation efficiency and reliability of the equipment through more effective planning of maintenance activities, thus improving the economic benefits of the enterprise.

In recent years, due to the development of big data and artificial intelligence technology, CBM has been widely concerned by the maintenance research community. In reference [5], a proportional risk model is used to describe the combined effects of aging and cumulative damage, and a CBM strategy is developed for systems susceptible to aging and cumulative damage. Nazanin *et al.* proposed a fuzzy early warning CBM strategy for automobile production line to provide early warning for potential production line failure or other dangerous situations, so as to make a more intelligent decision on maintenance strategy [6]. Phuc Do research group has established a binary system CBM model with stochastic and economic dependence. Taking the gearbox system as an example, in order to select one or more components for preventive maintenance, adaptive preventive maintenance and opportunistic maintenance rules are proposed [7]. In reference [8], a CBM strategy for soft and hard faults is proposed. Unlike traditional CBM strategies, this study does not use thresholds to define maintenance actions. The goal is to determine the optimal maintenance strategy to minimize the long-term expected average cost per unit time. The optimization problem is established in the framework of semi Markov decision process and solved by strategy iteration algorithm. In reference [9], a dynamic state maintenance model based on inverse Gaussian process is proposed for equipment maintenance decision-making with dynamic degradation. Firstly, an inverse Gaussian process with random parameters is proposed to describe the change of degradation characteristics of equipment during operation, and the important stochastic characteristics related to condition based maintenance are derived. Secondly, the dynamic maintenance threshold function is proposed, which can reduce the early failure risk of the equipment under the premise of ensuring a lower expected cost ratio. On this basis, a multi-objective dynamic maintenance decision model is established.

Most of the existing CBM related research is based on numerical calculation model, which requires input data for various operating conditions. Generally speaking, these parameter values need to be specially designed related sensors to detect data. From another point of view, based on the traditional equipment operation log information, this paper calculates the internal connection probability between various events to predict the possible failure of the equipment. The experimental results show that the method is feasible.

2. Event Based Prediction of Equipment Operation State

Whether it is the traditional manual record or the automatic generation of the system, the operation of the system will produce a large number of work logs recording all kinds of events during the operation of the system. Examples include hardware health, business processes, and server activity. The work log records all kinds of events occurred in the system, and these events must have certain internal relations. Therefore, it is theoretically feasible to predict the occurrence of specific events in the system through the event records in the work log.

2.1. Event Vector Model

Event vector model is used to calculate the probability model of a sequence of events. It is based on an event library.

Suppose $E = e_1^n \triangleq (e_1, e_2, \cdots, e_n)$ represents an event sequence composed of events in sequence, then the probability of the sequence events is

$$p(E) = p(e_1, e_2, \cdots, e_n)$$

Using Bayes formula, the above formula can be decomposed into

$$p(E) = p(e_1)\, p(e_2 \mid e_1)\, p(e_3 \mid e_1^2) \cdots p(e_n \mid e_1^{n-1}) \tag{1}$$

The (conditional) probability

$$\theta = \left(p(e_1), p(e_2 \mid e_1), p(e_3 \mid e_1^2), \cdots, p(e_n \mid e_1^{n-1})\right)$$

can be regarded as the parameters of the sequence event occurrence probability model. If these parameters are known, given a sequence event, the corresponding sequence event occurrence probability can be calculated quickly.

If the event $(e_1, e_2, \cdots, e_{n-1})$ has occurred, the probability of the event e_n is the probability of the sequence event *E*, $p(E)$.

The parameter θ is a vector of length *n*, which looks simple and its concrete calculation is not easy. If the number of possible atomic events is *m*, then the sequence events with length *n* have $C_m^n P_n^n$ possibilities. How to calculate these parameters efficiently is a problem that must be considered in practical application system.

2.2. Model Parameter Calculation

Considering the approximate calculation of $p(e_k \mid e_1^{k-1})(k > 1)$, using Bayes formula, we can get

$$p(e_k \mid e_1^{k-1}) = p(e_1^k) / p(e_1^{k-1})$$

According to the large number theorem, when the sample data is large enough, $p(e_k \mid e_1^{k-1})$ can be approximately expressed as:

$$p(e_k \mid e_1^{k-1}) \approx count(e_1^k) / conut(e_1^{k-1}) \tag{2}$$

where $count(e_1^k)$ and $count(e_1^{k-1})$ represent the number of occurrences of event sequences e_1^k and e_1^{k-1} in the sample, respectively.

It can be seen from formula (1): the probability of an event occurrence is related to all events before it. If we assume that the probability of an event is only related to the first *n*-1 events, that is

$$p(e_k \mid e_1^{k-1}) \approx p(e_k \mid e_{k-n+1}^{k-1})$$

Then formula (2) becomes:

$$p(e_k \mid e_1^{k-1}) \approx count(e_{k-n+1}^k) / count(e_{k-n+1}^{k-1}) \tag{3}$$

For formula (3), theoretically, the larger *n* is, the better the effect is. In fact,

DOI: 10.4236/jcc.2020.88004

when n is large to a certain extent, the improvement of model effect will be smaller. According to experience, n can be taken as 4.

There are two problems in formula (3):

1) If $count\left(e_{k-n+1}^{k}\right)=0$, it does not mean that $p\left(e_k \mid e_1^{k-1}\right)$ is 0;

2) If $count\left(e_{k-n+1}^{k}\right)=count\left(e_{k-m+1}^{k-1}\right)$, it does not mean that $p\left(e_k \mid e_1^{k-1}\right)$ value is 1;

In order to solve the above problems, a general method in machine learning field is used: after modeling the problem under consideration, an objective function is constructed first, and then the objective function is optimized to obtain a set of optimal parameters.

Considering the maximum log likelihood, the objective function can be designed as

$$L=\sum_{e\in S}\log p\left(e \mid PreEvent\left(e\right)\right) \tag{4}$$

where, S represents the event sample set, and $PreEvent\left(e\right)$ represents the previous event of event e.

In formula (4), when $PreEvent\left(e\right)$ is empty, $p\left(e \mid PreEvent\left(e\right)\right)=p\left(e\right)$ is taken.

According to formula (4), probability $p\left(e \mid PreEvent\left(e\right)\right)$ has been considered as a function of events e and $PreEvent\left(e\right)$, *i.e.*

$$p\left(e \mid PreEvent\left(e\right)\right)=F\left(e,PreEvent\left(e\right),\theta\right)$$

where θ is the selected parameter set. In this way, once the formula (2.4) is optimized to obtain the most optimized parameter set θ^*, F is only determined, and then the probability of event e can be calculated by $F\left(e,FreEvent\left(e\right),\theta^*\right)$ according to the previous event $PreEvent\left(e\right)$.

The calculation of the optimal parameters can be realized by the algorithm of neural probability language model (Word2Vec) [10].

Figure 1 shows the schematic diagram of the neural grid structure for the optimal parameters of formula (4), which consists of four layers: input layer, projection layer, hidden layer and output layer. Where W, u are the weight matrix between the projection layer and the hidden layer and between the hidden layer

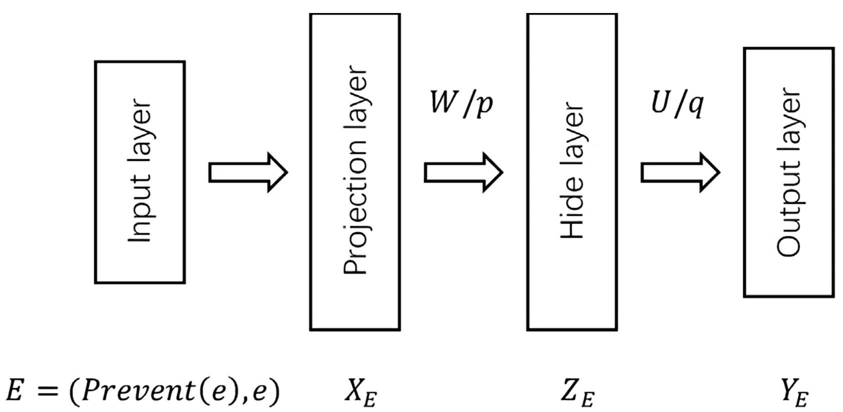

Figure 1. Schematic diagram of neural network structure.

and the output layer respectively, p, q are the offset vectors on the hidden layer and the output layer respectively.

For any event in the sample, $PreEvent(e)$ is taken as the previous $n-1$ events, so the binary $(PreEvent(e), e)$ is a training sample. If tanh function is used as the activation function of hidden layer, there are:

$$\begin{cases} Z_E = \tanh(WX_E + p) \\ Y_E = UZ_E + q \end{cases} \tag{5}$$

$Y_E = (y_1, y_2, \cdots, y_n)^{\mathrm{T}}$ calculated by formula (5) is only a vector of length n, and its component cannot represent probability. If you want the component of Y_E to represent the probability of event occurrence when the pre event is $PreEvent(e)$, you need to make a softmax normalization. After normalization, $p(e \mid PreEvent(e))$ can be expressed as

$$p(e \mid PreEvent(e)) = \frac{\exp(y_n)}{\sum_{i-1}^{n} \exp(y_i)} \tag{6}$$

3. Application Example

3.1. Data Set

In this study, 20,360 operation logs of a certain ship are selected for model training and testing. The operation log records a series of abnormal or parameter mutation conditions in chronological order. Table 1 shows an example of a diesel engine working log.

Event records only record some abnormal conditions, with different time intervals. Preprocess the event data as follows:

Take the event sequence according to the fixed time interval. If there is no event record in a certain time interval segment, fill with "normal" event.

Take 4 events successively from each event to form a sequence event sample.

If there are n events in a time interval segment S, the event of the time series is expanded to n sample sequences. When the sample sequence is expanded, the events before and after the S fragment are unchanged, and one of n events in the S fragment is taken to form an event sequence.

Table 1. Sample of event log.

Event	Time
Unstable speed	2018-05-10 13:30
Increased fuel consumption	2018-05-10 15:00
Exhaust pipe black smoke	2018-05-10 15:00
Exhaust pipe temperature too high	2018-05-10 15:00
Deformation of governor spring	2018-05-10 22:30
Cylinder block wear	2018-05-11 08:30
…	…

3.2. Analysis of Test Results

In model training, the probability of the last event is calculated according to the first three events of each sample. When the model is applied, refer to the preprocessing method of sample data. The first three events from the current time are taken as the previous events, and the fourth event is the possible event to be predicted, which is repeated and postponed. When the calculated probability of an event is significantly increased, the event is predicted to occur (not necessarily the probability is greater than a certain threshold).

Figure 2 records the effect of different total samples and events in a single sample on model training. It can be seen from the figure that the more the total number of samples, the higher the accuracy of the model. Under the condition that the total number of samples remains unchanged, the more events in a single sample, the higher the prediction accuracy. However, when the number of events in a single sample is greater than 4, the improvement of the prediction accuracy is not obvious. In fact, the occurrence of fault has a process from quantitative change to qualitative change. The prediction of a fault event needs to be based on multiple events in the time series. Generally speaking, the correlation of events far away from the time axis is relatively small, and the correlation between events far apart in sequence events can be ignored. Therefore, the possibility of using the first three events of sequence events to predict the occurrence of the fourth event according to the experiment is explainable. The number of events in a single sample is set as 4, that is, the first three events in each sample are preliminary events, and the last event is isometric events.

The results in Figure 3 show the probability changes of three events: "governor spring deformation", "fuel injection pump spring deformation" and "governor

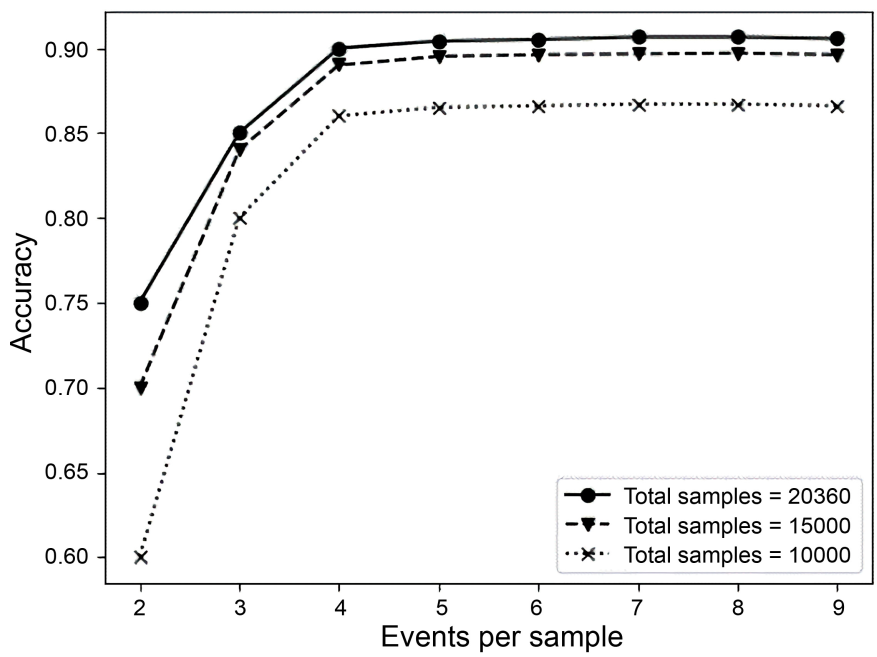

Figure 2. Accuracy of model training.

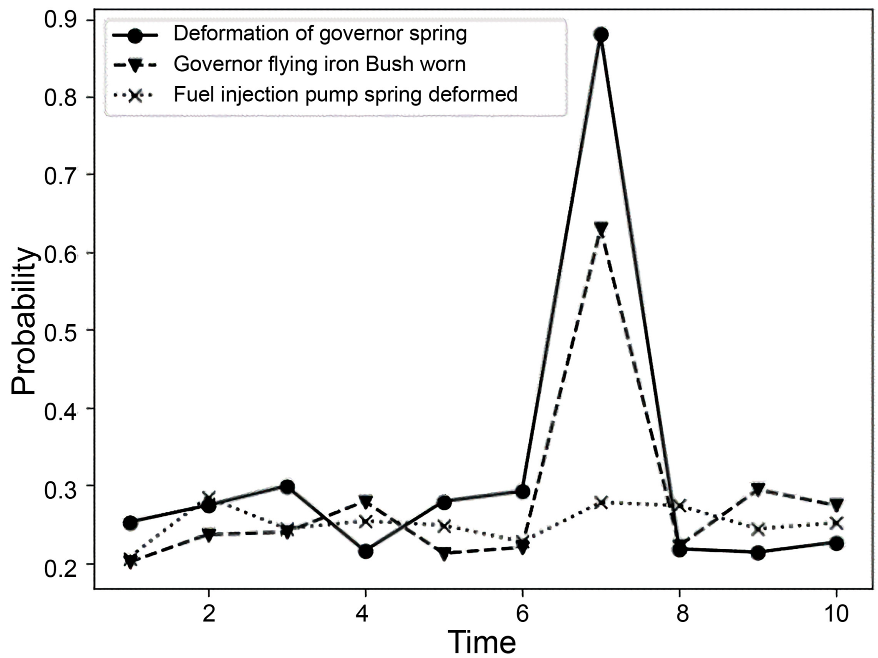

Figure 3. Predict the probability of the event.

flying iron Bush wear". Among them, the peak probability of "governor spring deformation" and "governor flying iron Bush wear" is 0.885 and 0.63 respectively, which indicates that it is more likely to occur. When the event probability of "governor spring deformation" increases obviously, the input event sequence is "oil consumption increases, normal and rotating speed is unstable". The predicted results show that "governor spring deformation" and "governor flying iron Bush wear" two events may occur, while "injection pump spring deformation" event is unlikely to occur. Moreover, according to the predicted results, the possibility of "regulator spring deformation" event is higher, so the governor spring should be checked first during maintenance.

In fact, when a diesel engine is running normally, as long as the working condition is certain, each cycle function of the diesel engine is the same, so its speed should also be stable. Only when the work of each cycle is different, the speed will be fast or slow. So there are three basic conditions to keep the speed of diesel engine stable:

1) The fuel supply of each cycle of the diesel engine is uniform (the quantity is the same) and the atomization quality of the fuel injector is uniform (the quality is the same).

2) Diesel engine governor is sensitive to regulation. Because the governor has the function of automatically adjusting the speed, that is, with the change of external load, automatically adjusting the fuel supply, so that the change of diesel engine speed is consistent.

3) The diesel engine has enough flywheel to store the inertia force and keep the engine running smoothly. This item is fully guaranteed by the general diesel engine manufacturer, and there will be no such problem in use.

There is no oil supply system problem in the predicted working condition (input event sequence), and the biggest problem basically occurs in the governor, so the prediction in **Figure 3** is in line with the actual situation.

4. Conclusions

The traditional planned maintenance and repair maintenance of equipment have been increasingly unable to keep up with the requirements of efficient modern production. In recent years, with the development of big data and artificial intelligence technology, CBM maintenance based on equipment performance and fault prediction has been widely concerned by the industry. Most of the mainstream equipment performance and fault prediction are based on various sensor data and mathematical models of the equipment. The complexity of the model and data signal is the key factor affecting the practicability of the model. In addition, even for the same type and batch of equipment, the manufacturing process, operation environment and other factors also affect the model parameters. In this study, the marine diesel engine is taken as the object, the event log of equipment operation is taken as the data source, and the probability of the target event is predicted by the sequential event occurrence model. Its data and goal are simple and direct, and more targeted. The experimental results show that the event model proposed in this paper is feasible.

This study attempts to use the historical data of log events to predict the feasibility of fault events. Our further work will be devoted to standardizing the equipment maintenance log and using various information sources to improve the practical performance of the research method.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Ziegel, E.R. (2004) System Reliability Theory: Models, Statistical Methods, and Applications. *Technometrics*, **46**, 495-496. https://doi.org/10.1198/tech.2004.s242

[2] Uhlmann, E., Stark, R., Rethmeier, M., Baumgarten, J., Bilz, M., Geisert, C., *et al.* (2015) Maintenance, Repair and Overhaul in Through-Life Engineering Services. *Through-Life Engineering Services*. Springer, Cham, 129-156. https://doi.org/10.1007/978-3-319-12111-6_9

[3] Li, Y., Tse, P.W., Yang, X., *et al.* (2010) EMD-Based Fault Diagnosis for Abnormal Clearance between Contacting Components in a Diesel Engine. *Mechanical Systems and Signal Processing*, **24**, 193-210. https://doi.org/10.1016/j.ymssp.2009.06.012

[4] Kumar, A., Shankar, R. and Thakur, L.S. (2018) A Big Data Driven Sustainable Manufacturing Framework for Condition-Based Maintenance Prediction. *Journal of Computational Science*, **27**, 428-439. https://doi.org/10.1016/j.jocs.2017.06.006

[5] Liu, B., Liang, Z., Parlikad, A.K., Xie, M. and Kuo, W. (2020) Condition-Based Maintenance for Systems with Aging and Cumulative Damage Based on Proportional Hazards Model. In: Crespo Márquez, A., Macchi, M. and Parlikad, A., Eds.,

Value Based and Intelligent Asset Management, Springer, Cham. https://doi.org/10.1007/978-3-030-20704-5_10

[6] Vafaei, N., Ribeiro, R.A. and Camarinha-Matos, L.M. (2019) Fuzzy Early Warning Systems for Condition Based Maintenance. *Computers & Industrial Engineering*, **128**, 736-746. https://doi.org/10.1016/j.cie.2018.12.056

[7] Do, P., Assaf, R., Scarf, P. and Iung, B. (2019) Modelling and Application of Condition-Based Maintenance for a Two-Component System with Stochastic and Economic Dependencies. *Reliability Engineering & System Safety*, **182**, 86-97. https://doi.org/10.1016/j.ress.2018.10.007

[8] Zheng, R. and Makis, V. (2020) Optimal Condition-Based Maintenance with General Repair and Two Dependent Failure Modes. *Computers & Industrial Engineering*, **141**, 106322. https://doi.org/10.1016/j.cie.2020.106322

[9] Wu, Z.Y., Guo, B., Axita, Tian, X. and Zhang, L. (2020) A Dynamic Condition-Based Maintenance Model Using Inverse Gaussian Process. *IEEE Access*, **8**, 104-117. https://doi.org/10.1109/ACCESS.2019.2958137

[10] Church, K.W. (2017) Word2Vec. *Natural Language Engineering*, **23**, 155-162. https://doi.org/10.1017/S1351324916000334