

Reinforcement Learning Based Predictive Maintenance for a Machine with Multiple Deteriorating Yield Levels[★]

Xiao WANG^{1,2}, Hongwei WANG¹, Chao QI^{1,*}, Appa Iyer Sivakumar³

¹*Institute of Systems Engineering/State Key Laboratory of Education Ministry for Image Processing and Intelligent Control, Huazhong University of Science and Technology, Wuhan 430074, China*

²*School of Safety Engineering, Shenyang Aerospace University, Shenyang 110136, China*

³*School of Mechanical and Aerospace Engineering/Division of Systems and Engineering Management, Nanyang Technological University, Singapore*

Abstract

This paper proposes a predictive maintenance methodology for a machine in manufacturing with deteriorating quality states represented by multiple deteriorating yield levels. Imperfect minor maintenance and perfect major repair are considered. The underlying yield level cannot be directly obtained. Instead, product quality inspection information is used as the observed system state. The optimal maintenance policy associated with each possible observed system state is learnt by modeling the problem as hidden semi-Markov decision processes and solving it using policy iteration based Q-P learning. Then the future maintenance time can be estimated by re-simulating the system model using the learned maintenance policy. A set of experimental studies is conducted to testify the effectiveness of the proposed methodology and to investigate the impacts of involved system parameters.

Keywords: Predictive Maintenance; Deteriorating Quality States; Multiple Yield Levels; Hidden Semi-Markov Decision Process; Q-P Learning

1 Introduction

Maintenance scheduling methodology has been evolving with the development of production and manufacturing technology, from corrective maintenance, through periodic preventive maintenance, to the recently prevalent condition-based maintenance. Condition-based maintenance has attracted considerable attentions from world-wide researchers and industrial practitioners with the main idea of determining whether maintenance action should be conducted under the current equipment state [1,2]. However, a question making much more sense from the practical perspective is that when the maintenance action should be conducted in the future if it is not necessary at present. Predictive maintenance is aiming at this question. With a predicted maintenance time, the production schedule can be adjusted in advance, so that the interruption on production caused by maintenance can be controlled.

*Project supported by the National Nature Science Foundation of China (No. 60904075).

*Corresponding author.

Email address: qichao@hust.edu.cn (Chao QI).

Researches on predictive maintenance are far from extensive in comparison with that on condition-based maintenance. Lu et al. [3] propose an approach to predict the machine deterioration condition for maintenance decision-making by adopting a state-space model and Kalman filtering technology. Wu et al. [4] provide a neural network model and used sensory information to predict the life percentage of the machine. Kaiser and Gebraeel [5] propose a sensory-updated degradation-based predictive maintenance policy. The sensory-updated scheme and a reliability-centered stopping rule are combined to decide maintenance schedule. Further, You et al. [6] develop a predictive maintenance method based on the policy proposed by Kaiser and Gebraeel [5] and a statistical lifetime distribution provided for long-term planning, by which the real-time residual life distribution of an individual system is predicted.

In most of the related works, the machine only has two possible states, operational or failure state, if without the influence of maintenance. The machine deterioration in these works means approaching to the end of lifecycle. In other words, operating a deteriorated machine has no impact on productivity performance. Under this assumption, the purpose of predictive maintenance is to predict the remaining lifetime or the machine failure time. However, in the real manufacturing systems, machine deterioration is closely related to productivity issues such as product quality. Before the machine fails to operate, it might experience a series of intermediate states, under which the machine is still operational but with a probability to produce conforming parts [7]. For instance, the machine might produce conforming parts with a higher probability when it is in a better condition in comparison with the situation when it deteriorates into a worse condition. That is, the machine has the character of multiple yield quality failures [8]. Here, yield is defined as the probability for the machine to produce a conforming part. For multiple yield quality failures there is no simple straightforward maintenance decision due to the tradeoff between effective productivity and maintenance cost.

This paper investigates a predictive maintenance methodology for a machine with multiple deteriorating quality states. Each quality state is directly related to a yield level. Besides the quality state deterioration process, the machine is influenced by severe malfunction as well, under which the machine fails to operate and cannot return back to an operational state without repair. Therefore, two types of maintenance actions: minor maintenance and major repair, are considered in this paper. Minor maintenance, which is imperfect maintenance, is purposefully initiated when the machine is operating on a deteriorated quality state. In contrast, the major repair, which is perfect maintenance, is compulsively triggered when the machine encounters a severe malfunction. The yield level associated with machine quality state cannot be directly obtained. Maintenance decision is made based upon the quality inspection information for the produced parts with inspection errors. The time parameters including production time per unit part and maintenance or repair time are generally distributed. A continuous-time, discrete-state hidden semi-Markov model is formulated for the optimal preventive of the machine. The maintenance policy corresponding to each observed state is learnt via policy iteration based reinforcement learning algorithm, under the condition of maximizing the long-run expected average reward rate of the system. Then the future maintenance time is estimated by re-simulating the system model using the learned maintenance policy.

2 Problem Description

The investigated system consists of a single machine with multiple deteriorating yield levels. A single part type is produced with a generally distributed production time and a fixed production

cost R_d . Each processed part is immediately inspected with zero inspection time and cost. Two types of inspection errors are considered. Type I error means there is a probability of e_1 to wrongly identify a conforming part as a defective part, associated with a cost of C_{e1} which includes production cost of one part and the opportunity cost for not selling a conforming part. Type II error means there is a probability of e_2 to wrongly identify a defective part as a conforming part, associated with a cost of C_{e2} which includes production cost of one part and other possible costs that far exceed production cost. The real quality level is identified by the current yield, which is the probability of producing a genuine conforming part. A genuine conforming part with correct inspection will immediately bring a fixed amount of benefit R_g to the system. Production cost R_d is the only expenditure associated with an accurately inspected defective part. Once a part is identified as a defective part, it will be immediately discarded without rework.

The yield level is the representative of the machine quality state. For each renewal cycle, the yield level starts from the best situation, e.g. 100%, and then gradually deteriorates to worse levels. There are multiple discretized yield levels, on each of them, the machine might sojourn for a generally distributed time period. Without intervention of maintenance, the machine will eventually fail to be operational due to severe malfunction. In this paper, this situation is encountered under either of the following two conditions: one is that the yield level reaches a prescribed lower limit; the other is that a random shock occurs with a stochastically distributed time to failure since last maintenance. Responding to a failure due to severe malfunction, a major repair action will be compulsively triggered with a fixed repair cost and a generally distributed repair time. The machine yield level will return back to the best state after the major repair action is completed. This kind of renewal cycles is illustrated by Fig. 1. For the case depicted by Fig. 1(a), the yield level starts to deteriorate from 100% until it reaches the lower limit W at t_1 , which results in the initiation of a major repair, with a fixed repair cost C_R considered. After the major repair is completed at t_2 , the machine returns back to its best yield level, and starts another renewal process. For the case illustrated by Fig. 1(b), the major repair is initiated by the random shock occurs on t_1 .

When the machine is operating with a deteriorated yield level, minor maintenance can be purposefully triggered associated with a fixed maintenance cost C_M and a generally distributed maintenance time. This kind of minor maintenance is imperfect, which means that the machine cannot return back to its best state [9]. In other words, within a renewal cycle, i.e. the time period from the time when the machine yield starts to deteriorate from its best level to the time when the yield returns back to its best level, minor maintenance might be conducted for a couple of times. This is illustrated by Fig. 1(c). Within the example of depicted renewal cycle, there are three sub-cycles. For the first sub-cycle $k = 1$, the yield level starts at y_{11} , and then experiences another three yield levels, y_{12} , y_{13} , and y_{14} . A minor maintenance is triggered at t_1 . After the minor maintenance, the yield level returns back to the level y_{21} , which is between y_{11} and y_{14} , and starts the second sub-cycle $k = 2$. This renewal cycle is ended after a major repair is initiated at t_5 and completed at t_6 . Here, y_{kl} denotes the l th yield level at k th sub-cycle, and λ_{kl} denotes the sojourn time of y_{kl} .

For the practical perspective, this paper considers the situation when the underlying yield level cannot be directly obtained. The system quality state can only be observed by product quality inspection information. There are a variety types of quality information to be adopted as observed system state, such as the inspection results of the consecutively produced parts within a given time window. In this paper, we adopt (p, b) , where p is the total number of parts produced since last maintenance or repair, and b is the number of defective parts out of p . The decision-making

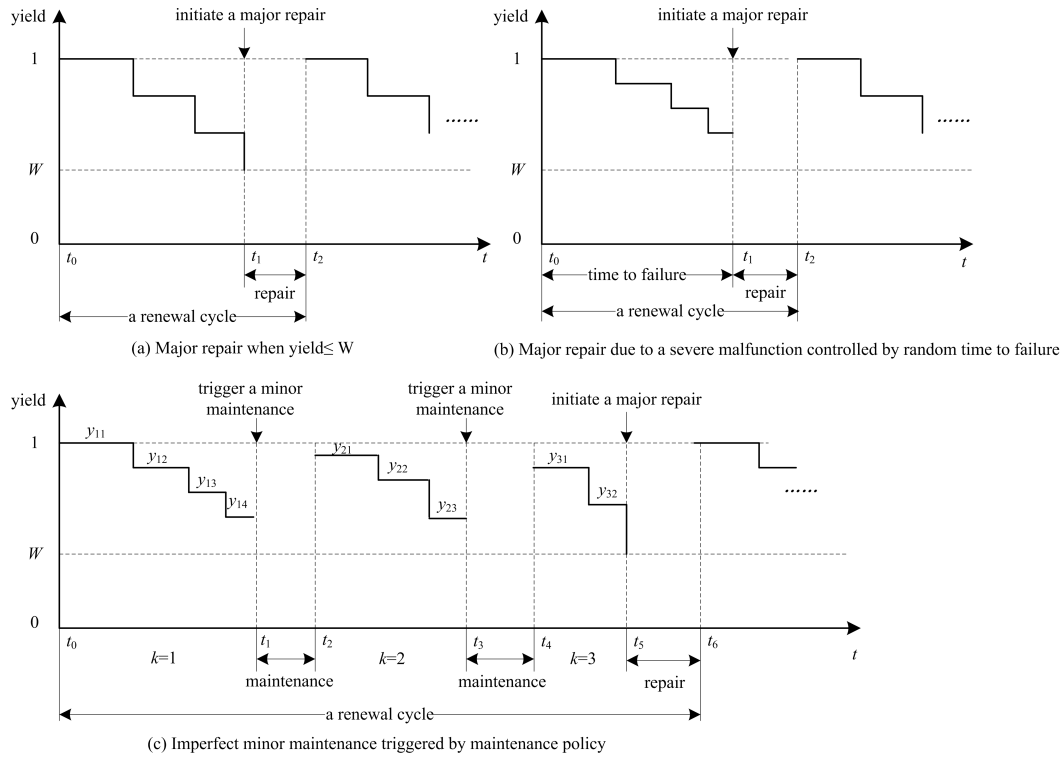


Fig. 1: Illustration of quality state renewal cycles

problem is when the minor maintenance action should be conducted for a given state $x = (k, p, b)$ so that the system average reward can be maximized over a long time horizon.

3 Reinforcement Learning Based Predictive Maintenance

3.1 Hidden semi-Markov decision process model

A hidden semi-Markov decision process model is utilized to represent the deteriorating process of the machine, due to the generally distributed sojourn time and the unobservable underlying stochastic process on deteriorating quality states. The decision points in this process are the time points when a new part is produced and inspected. The underlying yield level y_{kl} cannot be directly obtained. Instead the observed system state is the product quality inspection information at the k -th sub-cycle, $x = (k, p, b)$, at which the action space consists of two possible actions, $a \in A\{1, 0\}$. When $a = 1$, the machine continues to produce with a continuously deteriorating underlying yield. When $a = 0$, the machine is stopped to conduct a minor maintenance, after which the machine returns back to a better but not the best yield level, and another $(k+1)$ -th deterioration sub-cycle is started. At each observed system state, when the yield reaches a low limit or when a severe random shock is encountered, a major repair might be compulsively initiated, which will bring the machine back to its best yield level and start another renewal cycle.

The sojourn time λ_{kl} of each yield level y_{kl} can be generally distributed, and here it is assumed to follow a gamma distribution, $\lambda \sim \Gamma(\alpha_{kl}, \beta)$. Also, the sojourn time λ_{kl} decreases with the increase of l as $\alpha_{k,l+1} = b_s \alpha_{kl} (0 < b_s < 1)$. The time to random failure in k th sub-cycle follows a gamma distribution with shape parameter α_k and scale parameter β , and $\alpha_k = \sum_{l=1}^L \alpha_{kl}$, the expectation

of which is gradually shortened with machine deterioration constrained by the following equation

$$\alpha_{k+1} = y_{k+1,1}\alpha_k \quad k = 1, 2, \dots, K, \quad l = 1, 2, \dots, L \quad (1)$$

3.2 The Q-P learning algorithm

Since the value iteration is inexact for solving average reward hidden semi-Markov decision processes model [10], the policy iteration based reinforcement learning algorithm, Q-P learning algorithm is adopted for solving the model. The Q-factors updating equation can be expressed as follows:

$$Q(x, a) \leftarrow (1 - \alpha)Q(x, a) + \alpha[r(x, a, x') - \rho t(x, a, x') + Q(x', \arg \max_{h \in \{0,1\}} P(x', h))] \quad (2)$$

where $r(x, a, x')$ is the actual cumulative reward between two successive decision epochs, $t(x, a, x')$ is the actual sojourn time between the decision epochs, α is a learning rate parameter, and P-factors show the initial policy π . For the system studied in this paper, we have:

$$r(x, a, x') = \begin{cases} R_g & \text{when a conforming part is accurately inspected,} \\ -C_{e1} & \text{when a conforming part is wrongly inspected with type I error,} \\ -R_d & \text{when a defective part is accurately inspected,} \\ -C_{e2} & \text{when a defective part is wrongly inspected with type II error,} \\ -C_R & \text{when a major repair is compulsively constructed,} \\ -C_M & \text{when a minor maintenance is purposefully triggered} \end{cases}$$

$\alpha = \alpha_0/(1 + u_v)$ with $\alpha_0 = 0.1$, $u_v = V(x, a)^2/(N_{max} + V(x, a) - 1)$, N_{max} is large number, and $V(x, a)$ denotes visit-times for state-action pair (x, a) .

The maintenance policy for different observed system states can be learnt by simulating the deteriorating model of the machine and incorporating the learning Eq. (2).

3.3 The optimal predictive maintenance time

With the obtained maintenance policy, the predictive maintenance time for each observed system state x can be estimated by re-simulating the deteriorating process model of the machine. At any time t as the simulation process enters a state x , the action a corresponding to the maintenance policy provides the optimal prediction choice for the state, and then the future maintenance time associated with the state x can be derived. For instance, at any time point t_0 , the observed system state is x_i , and the action $a = \pi^*(x_i)$ is taken. Under the influence of the action a , the next decision-making state x_j is obtained. This sequence repeats till the termination condition is called, namely, minor maintenance action occurs at time point t_1 . Then the predictive maintenance time for the observation state x_i can be obtained from the result of $(t_1 - t_0)$. The process is different from the content of the above section because it does not involve any learning. We calculate the predictive maintenance time from several replications and then declare the mean of the values

obtained from all the replications to be the final predictive maintenance time. This mean value represents a “good” estimate of the predictive maintenance time.

Over all, the methodology for obtaining the predictive maintenance time includes three critical steps: policy evaluation, policy improvement [10], and predictive maintenance time estimation, as described below:

(1) In the policy evaluation step, a random policy π is initialized by randomly setting the values of $P(x, a)$. The average reward rate of the policy ρ , is estimated by running the deteriorating model for a long enough time using the policy. Whenever the system enters a new state x' from state x under the influence of action a , the immediate reward of state transition $r(x, a, x')$ and the transition time $t(x, a, x')$ are calculated, based on which the Q-factor $Q(x, a)$ is updated by equation Eq. (2). Subsequently, the condition to compulsively conduct major repair is checked, either when the yield level falls below the prescribed level ($y_{kl} \leq W$), or when a severe malfunction is encountered, i.e. when the accumulated time since last repair/maintenance has exceeded a randomly generated time to failure ($T_c \geq T_f$). If a major repair is not conducted, a minor maintenance decision will be made, by choosing production ($a = 1$) or maintenance ($a = 0$) with equal probability. If a major repair is conducted, the iteration number N will be updated by increment of 1. When N reaches a prescribed large number N_{max} , the process will get into policy improvement phase.

(2) In the policy improvement step, the P-factors are updated by $P(x, a) \leftarrow Q(x, a)$, and the number of episode E is updated by increment of 1. The new average reward rate ρ is estimated according to the new policy from P-factors. When $E \geq E_{max}$, the optimal maintenance policy π^* can be defined by $\pi^*(x) = \arg \max_{a \in A} P(x, a)$.

(3) In the predictive maintenance time estimation step, the valid production time is accumulated from any given state x_i to minor maintenance state, and it is perceived as predictive maintenance time $T_M^j(x_i)$. In the process, the visit-times n and the sum of $T_M^j(x_i)$ for each observed state x_i are recorded, and then the mean value $\frac{1}{n} \sum_{j=1}^n T_M^j(x_i)$ is declared as the final estimation of the predictive maintenance time.

4 Simulation Experiments

4.1 Yield deterioration process design

The yield deterioration process designed in this simulation experiment is referred to that described in Zhu et al. [11, 12]. For the two consecutive deterioration sub-cycles, the continuous yield functions have the following relationship:

$$Yield_{k+1}(t) = b_k Yield_k(t + a_k D_k) \quad t \in (0, D_{k+1}) \quad (3)$$

Where b_k is the yield decrease factor which makes a decrease in the yield function, and $0 < b_k \leq 1$; a_k is the age reduction factor which makes non-one yield after maintenance, and $0 \leq a_k < 1$; D_k is the time duration of sub-cycle k . When $t = 0$, $Yield_{k+1}(0)$ denotes where the yield level returns to after the k -th minor maintenance action. In this study, the linear deteriorating mode is assumed as follows:

$$Yield_1(t) = 1 - k_y t \quad (4)$$

where $k_y = 1/200$. The discrete yield level is decided by the following equation, and L is a prescribed number of yield levels for each k .

$$y_{k+1,l} = Yiled_{k+1}(0)(1 - \frac{l-1}{L-1}) \quad l = 1, 2, \dots, L \tag{5}$$

4.2 Base-line case

In the base-line case below, we assume that the production time of each part follows a gamma distribution $\Gamma(10, 0.1)$; the minor maintenance time follows a uniform distribution $U(2, 8)$; the major repair time follows a gamma distribution $\Gamma(20, 0.5)$; the time to random shock failure when $k = 1$ follows a gamma distribution $\Gamma(100, 0.2)$. The critical yield level W is set to be 0.6, and L is set to be 4. Other base-line case parameters are given in Table 1.

Table 1: Base-line case parameters

a_k	b_k	b_s	e_1	e_2	C_{e1}	C_{e2}	R_g	R_d	C_M	C_R
0.2	0.9	0.8	0.05	0.05	30	60	90	40	100	300

As shown in Fig. 2, with the increasing number of the policy improvement, the average reward rate converges to a steady value 26.5.

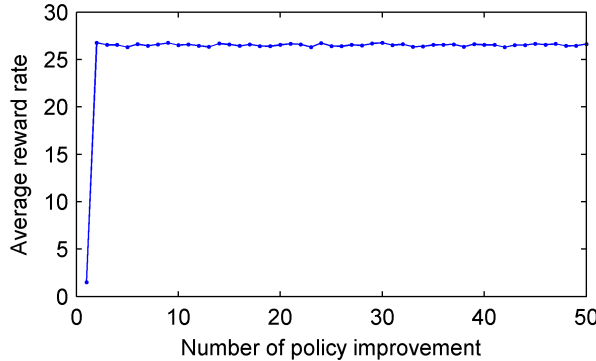


Fig. 2: Convergence of average reward rate

Fig. 3 illustrates the obtained maintenance policies when k switching from 1 to 5, respectively. The horizontal axis represents the total number of produced parts p since last maintenance/repair, and the vertical axis represents the number of defective parts b out of p . As illustrated in Fig. 3, when the number of sub-cycle k becomes larger, the probability of producing defective parts will be higher for fixed value p , and the maintenance actions are performed earlier. Above phenomena demonstrates that the maintenance decision rule becomes more conservative with the increasing number of sub-cycles and defective rate.

The predictive maintenance time can be produced based on the above maintenance policy. In Fig. 4, the horizontal axis represents the total number of the parts p , and the vertical axis represents the predictive maintenance time. It can be observed that the time for taking the machine down depends on observation state (k, p, b) . It decreases as p increases, and also it decreases monotonously as b increases under a given value of p . Moreover, the increasing sub-cycle k gives rise to the decreasing of predictive maintenance time.

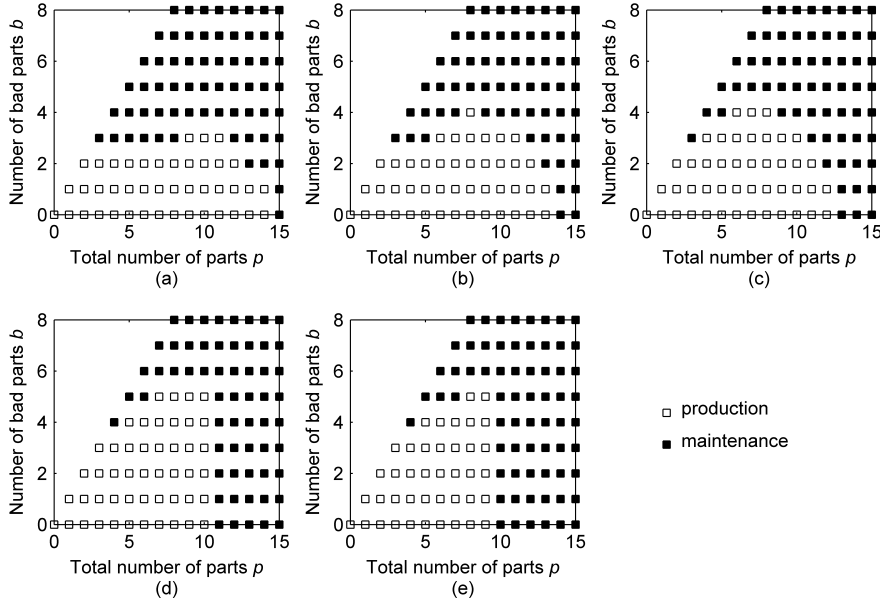


Fig. 3: Maintenance policy from $k=1$ to $k=5$

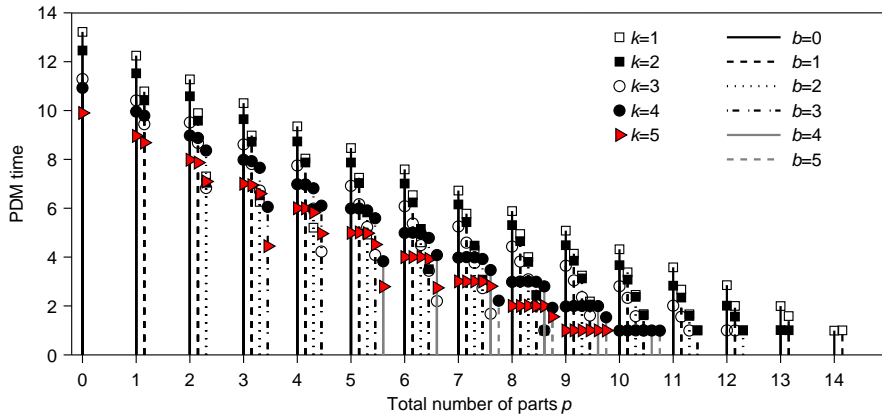


Fig. 4: Predictive maintenance time from $k=1$ to $k=5$

4.3 Impact of yield decrease factor

The yield decrease factor b_k influences the deterioration rate of the yield function. As shown in Fig. 5, when b_k decreases, the predictive maintenance time decreases obviously. For observation state (3, 4, 1), for example, the decrement of predictive maintenance time varies from 0.5 to 1.99 when b_k varies from 0.95 to 0.8, because the machine is restored to a worse quality state after the maintenance when b_k is smaller. When b_k is equal to 0.8, the predictive maintenance barely changes with the variation of b_k , because of the lower quality level of the machine.

4.4 Impact of the profit of conforming productivity

As shown in Fig. 6, the increasing of the profit of conforming productivity R_g leads to an increase of predictive maintenance time. The increment of predictive maintenance time tends to

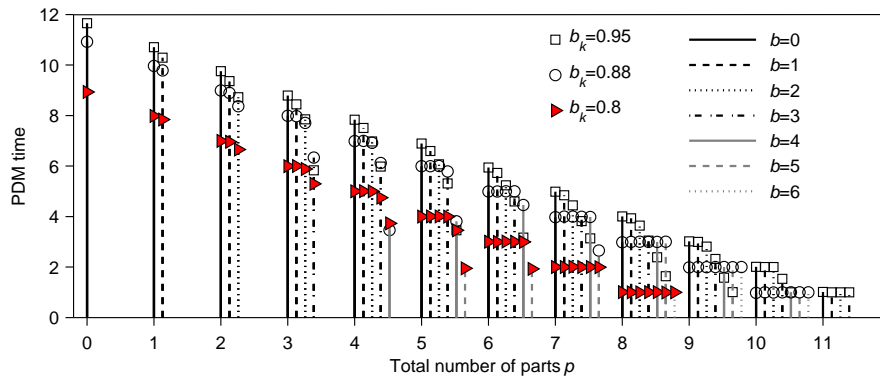


Fig. 5: Impact of yield decrease factors

a downward trend from 0.74 to 0.13 for the same observation state (2, 4, 1). The reason is that the long-run expected average cost rate is increased, and it varies from 11.76 to 66.58 with the increasing of R_g . Therefore, the predictive maintenance time is postponed.

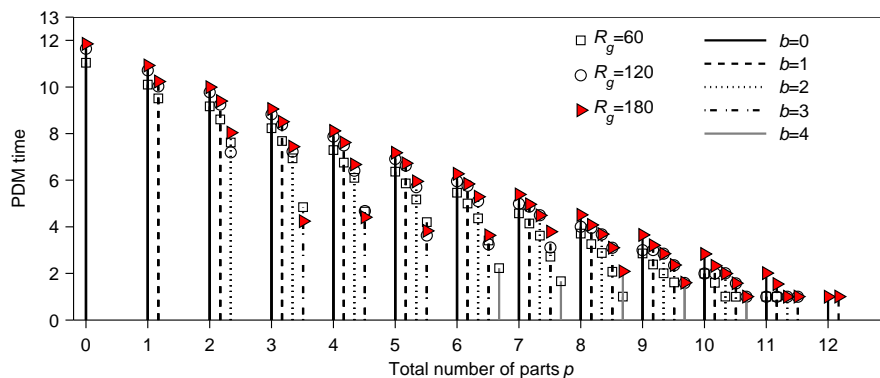


Fig. 6: Impact of profits of conforming productivity

4.5 Impact of the cost of defective productivity

As shown in Fig. 7, the decrement of predictive maintenance time tends to a downward trend as the increasing of the cost of defective productivity R_d , such as for the same observation state (2, 4, 1), the decrement is 1.35 and 0.32. The reason is that the decision maker wants to avoid that the system gets older and produces defective parts with a higher probability. As a result, the machine is taken down to minimize the long-run expected average cost rate.

4.6 Impact of type I error

As shown in Fig. 8, as the probability of Type I error e_1 increases, there is a slight decrease on the predictive maintenance time for each observation state. For example, given the observation state (1, 4, 1), the decrement is 0.43 and 0.21 respectively when e_1 varies from 0 to 0.1. There are two reasons for this. On the one hand, the increasing of e_1 leads to a loss of long run expected average profit rate, and it decreases from 31.8 to 24.7, and on the other hand, the cost resulting from

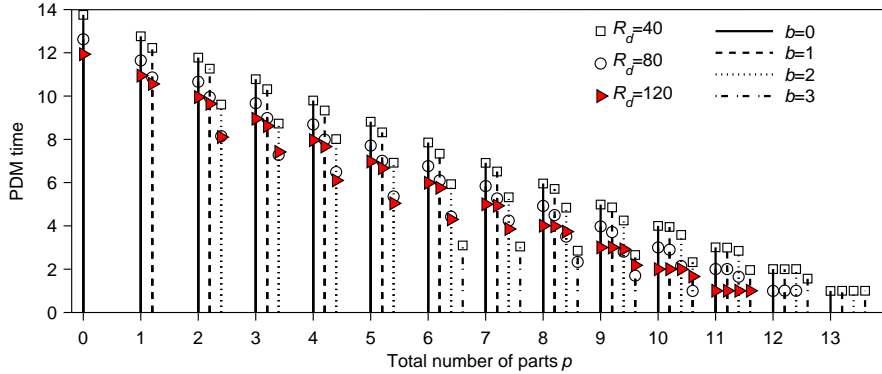


Fig. 7: Impact of costs of defective productivity

Type I error $C_{e1}=40$ is relatively small. Similar to the impact of Type I error, a slight decrease in predictive maintenance time with the increasing of the probability of Type II error e_2 can be incurred, since the cost of Type II error $C_{e2}=100$ is relatively small.

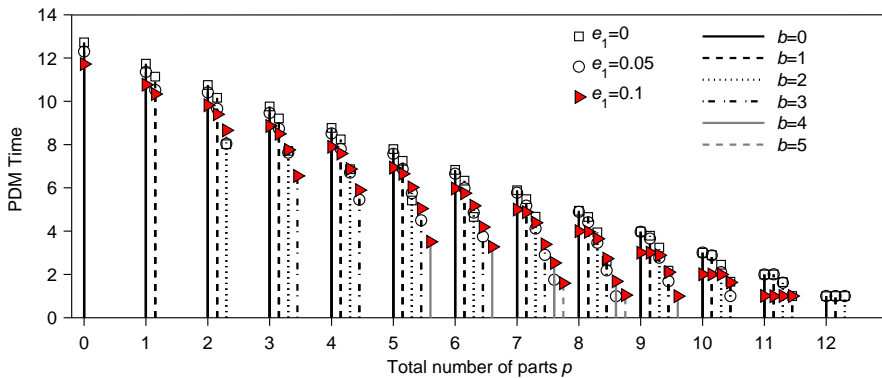


Fig. 8: Impact of type I error

5 Conclusion

This paper has focused on the development of a predictive maintenance methodology for a single machine with deteriorating quality states represented by multiple yield levels. The underlying yield level cannot be directly observed, and the product quality inspection information is used as system observation state. The production, minor maintenance and major repair times are considered to be generally distributed. The hidden semi-Markov decision processes model is utilized to present the deteriorating processes of the machine. The optimal maintenance policy is obtained by solving the model using Q-P learning algorithm. The future maintenance time associated with each observed state is estimated by re-simulating the deteriorating processes model using the learned maintenance policy. The results of simulation experiments indicate that the predictive maintenance methodology based on reinforcement learning possesses potential capability to adapt to the dynamics in manufacturing systems.

Acknowledgement

This work was supported by the National Science Foundation of China (No. 60904075).

References

- [1] A. Ghasemi, S. Yacout, M. S. Ouali. Optimal condition based maintenance with imperfect information and the proportional hazards model. *International Journal of Production Research*, 45 (4), 2007, 989-1012.
- [2] M. L. Neves, L. P. Santiago, C. A. Maia. A condition-based maintenance policy and input parameters estimation for deteriorating systems under periodic inspection. *Computers and Industrial Engineering*, 61 (3), 2011, 503-511.
- [3] S. Lu, Y. Tu, H. Lu. Predictive condition-based maintenance for continuously deteriorating systems. *Quality and Reliability Engineering International*, 23 (1), 2007, 71-81.
- [4] S. Wu, N. Gebraeel, M. Lawley, Y. Yih. A neural network integrated decision support system for condition-based optimal predictive maintenance policy. *IEEE Transactions on Systems, Man and Cybernetics-Part A: Systems and Humans*, 37 (2), 2007, 226-236.
- [5] K. A. Kaiser, N. Z. Gebraeel. Predictive maintenance management using sensor-based degradation models. *IEEE Transactions on Systems, Man, and Cybernetics- Part A: Systems and Humans*, 39 (4), 2009, 840-849.
- [6] M. Y. You, F. Liu, W. Wang, G. Meng. Statistically planned and individually improved predictive maintenance management for continuously monitored degrading systems. *IEEE Transactions on Reliability*, 59 (4), 2012, 744-753.
- [7] S. A. Mosley, T. Teyner, R. M. Uzsoy. Maintenance scheduling and staffing policies in a wafer fabrication facility. *IEEE Transactions on Semiconductor Manufacturing*, 11 (2), 1998, 316-323.
- [8] I. C. Schick, S. B. Gershwin, J. Kim. Quality/quantity modeling and analysis of production lines subject to uncertainty. Report, Laboratory for Manufacturing and Productivity, Massachusetts Institute of Technology, 2005.
- [9] H. Pham, H. Wang. Imperfect maintenance. *European Journal of Operational Research*, 94 (3), 1996, 425-438.
- [10] A. Gosavi. *Simulation-based optimization: parametric optimization techniques and reinforcement learning*. Norwell: Kluwer Academic Publishers, 2003.
- [11] H. Zhu, F. Liu, X. Shao, Q. Liu, Y. Deng. A cost-based selective maintenance decision-making method for machining line. *Quality and Reliability Engineering International*, 27 (2), 2011, 191-201.
- [12] L. Hu, D. Yue. Probabilistic analysis of series system with M failure modes and a repairman providing services to customers. *Journal of Computational Information Systems*, 6 (5), 2010, 1495-1500.