Journal of the KSQC
Vol. 16, No.2 1988

Machine Maintenance Policy Using Partially Observable Markov Decision Process

Pak, Pyoung Ki*
Kim, Dong Won*
Jeong, Byung Ho**

ABSTRACT

This paper considers a machine maintenance problem. The machine's condition is partially known by observing the machine's output products. This problem is formulated as an infinite horizon partially observable Markov decison process to find an optimal maintenance policy. However, even though the optimal policy of the model exists, finding the optimal policy is very time consuming. Thus, the intends of this study is to find ε −optimal stationary policy minimizing the expected discounted total cost of the system. ε −optimal policy is found by using a modified version of the well−known policy iteration algorithm. A numerical example is also shown.

1. Introduction

This paper deals with a machine maintenance problem with some internal components. The components are assumed to be identical. In this situation, the machine must be dismantled to know the status of these components. That is, the true condition of the machine can be known completely with only costly machine inspection. The machine's condition can be partially known by observing the machine's output products. Thus, this partial information for the

* Cheonbuk National University
** Korea Advanced Institute of Science and Technology

machine's condition can be used to control the system.

Thus, this problem can be formulated as a well-known partially observable Markov decision process(POMDP). POMDP is a generalization of Markow decision process(MDP) that permits the state uncertainty of the system, and that allows the acquisition of the partial information for the current state. (Albright(1979), Platzman(1980), Sondik(1978)) The partial information for the current state is obtained by examining the output products of the machine at lower cost than cost of inspecting the machine.

The commonly proposed methods for solving the infinite horizon POMDP are policy iteration algorithm(Sondik(1978)) and successive approximation method. The former converges faster than the later but value determination step in policy iteration algorithm has severe computational burden. Sawaki(1980) provided the algorithm which applies the successive approximation method to the value determination step of the policy iteration algorithm. However, since the approximated expected cost of a stationary policy is piecewise linear but, in general, not concave, one-pass algorithm(Smallwood and Sondik(1973)) can not be applied to policy improvement step. Thus, this study substitutes the concave function sufficiently close to the approximated expected cost of a stationary policy for the expected cost of the stationary policy.

2. Model Description

Consider a machine maintenance problem with n internal components, each of which must operate on the product before it is finished. Assume that the components are assumed to be identical, then, the internal dynamics of the machine can be modeled as a $(n+1)$ state discrete time Markov process with the $(n+1)$ states corresponding to zero, one, $\cdots n$ identical components that have failed. That is, let $s(t)$, which means the number of failure components at time t, be a random variable defined on a sample space, $\Omega=\{0, 1, 2, \cdots, n\}$, where discrete time $t \in I$, $I=1, 2, \cdots$If component breakdowns are independent of each other, and if there is a probability that an operational component will breakdown during the manufacture of a product, then a transition matrix is constructed using the probability. Thus, the stochastic process $\{s(t), t \in I\}$, called the core process or the underlying process, can be assumed to be a finite state Markov chain with stationary $(n+1) \times (n+1)$ transition probability matrix $P=(P_{ij})$, $i, j \in \Omega$.

Then, the core process represents the condition of the machine which is deteriorating over time. However, in many real system, the true condition of the machine is not known with certainty. Even if the true condition can be observed perfectly of directly, it incurs high inspection cost(e.g., opportunity cost due to system's interruption). Thus, the partial information about the state of the machine obtained by observing the machine's products is used to control the system efficiently.

For this maintenance problem, there are four control alternatives available during each production cycle. First, we simply manufacture another item, but without examining whether the resulting item is defective or not($k=1$). For the second alternative, we proceed as in manufacture alternative, except that the finished product is examined at a sampling cost($k=2$).

There are two observable outputs for this alternative corresponding to the production of a nondefective or defective product. Third, the manufacturing process is interrupted for one production cycle, the machine is dismantled and the n internal components are inspected and replaced if they have failed($k=3$). The third alternative incurs the replacement cost and additional inspection cost. The last control alternative involves the replacement of n internal components with no prior inspection($k=4$). This alternative accrues the replacement cost for all internal component, but does not incur an inspection cost.

The maintenance problem is formulated as infinite horizon POMDP. Then, the core process is completely described by p^k and the initial state vector,

$$\pi(0)=\{\pi_0(0),\ \cdots,\ \pi_n(0)\},\quad \text{where } \pi_i(0)=Pr\{s(0)=i\}$$

A partially observed state for the current state is denoted by θ. The probabilistic characteristics of this current partial observation is represented by stockastic matrices $B=[b_{i\theta}^k]$, where $b_{i\theta}^k$ represents the probability that θ is observed under action k when the machine is in state i. Smallwood and Sondik(1973) gives the information structure of this model and provides the rule of calculating the state vector $\pi(t)$ based on the previous state vector $\pi(t-1)$ and partial observation θ.

$$\pi_j(t)=\frac{\sum_i\pi_i(t-1)p_{ij}^kb_{j\theta}^k}{\sum_{ij}\pi_i(t-1)p_{ij}^kb_{j\theta}^k}$$

This transformation function can be written is terms of vectors and matrices as follows :

$$\{\theta|\pi,k\}=\pi P^kB_\theta^k\underline{1}$$

and

$$T(\pi|k,\theta)=\frac{\pi P^kB_\theta^k}{\{\theta|\pi,k\}} \tag{1}$$

where $\underline{1}=[1,1,\cdots,1]^T$, $B_\theta^k=\text{diag}[b_{i\theta}^k]$. The superscript T denotes transpose of a vector. $\{\theta|\pi,k\}$ represents the probability that the observed state after one transition is θ, when the current state vector is π and action k is chosen. These functions define the dynamics of the state of knowledge of the core process. Suppose that the sole source of information from the process is the sequence of observations, and that the state vector $\pi(t)$ is computed after the observed state is revealed. Then, the state vector is a sufficient statistics of the state of knowledge (Smallwood and Sondik(1973)).

Denote the immediate cost operating the process at the state i under action k by q_i^k. For example, the immediate cost of the third alternative at state is inspection cost plus i times the replacement cost per unit component plus additional inspection cost. When the state vector is π and action k is chosen, then the expected immediate cost $\pi q^k=\sum_i\pi_iq_i^k$ is incurred to operate the process. Let β be a discount factor, $0\leq\beta<1$. Denote $\delta_t(\pi)$ as a control function that represents the action to be chosen when the state vector at time t is $\pi(t)$. Then, the expected discounted cost control problem for fixed $\pi(0)$ can be written as.

$$\underset{\delta 0,\delta 1 \cdots}{\mathrm{MIN}}\ E\left[\sum_{t=0}^{\infty} \beta^t \pi(t) q^{\delta t(\pi)}\right] \tag{2}$$

where the sequence of control functions $(\delta_1,\ \delta_2,\cdots)$ must be selected to minimize the expression. Such a sequence of control functions is defined as a policy. If a policy consists of a single control function to be used at each time period, then the policy is termed stationary. A stationary policy is denoted by $\delta^\infty=(\delta,\ \delta\cdots)$. Then, the problem (2) is within the scope of Blackwell's formulaton. (1965). It follows that the minimum expected discounted cost as a function of the initial state vector π, $C^*(\pi)$, exists and that it satisfies

$$C^*(\pi)=\mathrm{MIN}_k[\pi q^k+\beta\Sigma_\theta\{\theta|\pi,k\}\ C^*(T(\pi|\theta,\ k))] \tag{3}$$

Furthermore, there exists a stationary policy that achieves this minimum cost. and the stationary policy is denoted by $(\delta^*)^\infty$, where $\delta^*(\pi)$ is the minimizing alternative in (3). Thus, Eq(3) can be written as

$$C^*(\pi)=\pi q^{\delta^*}+\beta\Sigma_\theta\{\theta|\pi,\ \delta^*\}C^*(T(\pi|\theta,\delta^*)) \tag{4}$$

where the specific dependence *of* $\delta^*(\pi)$ on π has been suppressed.

Now, consider a stationary policy, δ^∞. Let $C(\pi|\delta)$ be the expected discounted cost of a stationary policy, δ^∞, at an initial state vector π. Then, it is well known that $C(\pi|\delta)$ is the unique bounded solution to *Eq.* (5). (Blackwell(1965))

$$C(\pi|\delta)=\pi q^\delta+\beta\Sigma_\theta\{\theta|\pi,\delta\}\ C(T(\pi|\theta,\delta)|\delta) \tag{5}$$

However, since it is far from trivial to find the solution of(5), a method to approximate the expected discounted cost for a stationary policy is needed. The following section presents some properties of $C(\pi|\delta)$ in(5) and an approximation method of $C(\pi|\delta)$ based on these properties. In the latter part of this paper, for notational simplicity, an approximation of $C(\pi|\delta)$ is denoted by $f^m(\pi)$, where m is the iteration number and δ is suppressed.

3. The Approximation of $C(\pi|\delta)$

In this section, the expected discounted cost of any stationary policy is approximated by using successive approximation method. The approximated expected discounted cost of the stationary policy can be used in value determination step of the algorithm to find ϵ-optimal policy. This idea is originally proposed by Sawaki(1980). Define the following operators $U_k f$, $U_\delta f$ and $U_* f$ on any real valued bounded function f.

$$(U_k f)(\pi)=\pi q^k+\beta\Sigma_\theta\{\theta|\pi,\ k\}f(T(\pi|\theta,\ k))\} \tag{6}$$

$$(U_\delta f)(\pi)=\pi q^k+\beta\Sigma_\theta\{\theta|\pi,\ \delta\}f(T(\pi|\theta,\ \delta))\} \tag{7}$$

$$(U_* f)(\pi)=\mathrm{MIN}_k[\pi q^k+\beta\Sigma_\theta\{\theta|\pi,\ k\}f(T(\pi|\theta,\ k))] \tag{8}$$

Then, since the above operators satisfy the monotone contraction assumption, they are monotone contraction mappings. Some important properties given by Sawaki are reviewed without proof.

Theorem 1. Suppose that $f(\pi)$ is a piecwise linear concave function on Π. Then, the operator $(U_k f)(\pi)$, is also a piecewise linear concave function on π for each k.

Theorem 2. Suppose that $f(\pi)$ is piecewise linear. Then,

i) $(U_\delta f)(\pi)$ is piecewise linear whenever δ is a simple policy.

ii) $(U_* f)(\pi)$ is piecewise linear concave and there exists a simple policy δ such that $U_\delta f = U_* f$

The above two theorems are proved by Sawaki(1980). Theorem 2 plays the fundamental role of the algorithm to be presented in the next section. That is, the part i) of theorem 2 implies that the expected discounted cost of a stationary policy δ can be approximately obtained by applying the operator $U_\delta f$ repeatedly, and that the expected cost is piecewise linear on Π. Furthermore, the part ii), in fact, indicates the policy improvement step can be performed by using one-pass algorithm. Since the part i), however, does not guarantee the concavity of the expected discounted cost of a stationary policy δ, the algorithm can not be directly applied to policy improvement. Thus, the concave hull, defined as $\bar{f}(\pi)=\min_i\ [\pi\alpha_i]$ by Sondik(1978), of a piecewise linear function, $f(\pi)=\pi\alpha_i$ for $\pi\in E_i$, will be used in the policy improvement step.

Corollary. Suppose that $f^1(\pi)$ is piecewise linear on Π and $\bar{f}^1(\pi)$ is the concave hull of $f^1(\pi)$. Then,

i) $f^m(\pi)=(U_\delta f^{m-1})(\pi)$, $m=1,\ 2,\ \cdots$, is piecewise linear for any simple policy δ.

ii) $f^m(\pi)=(U_* f^{m-1})(\pi)$, $m=1,\ 2,\ \cdots$, is piecewise linear concave function on Π and there exists a simple policy, δ^m, satisfying $(U_{\delta m}\bar{f}^{m-1})(\pi)=(U_*\bar{f}^{m-1})(\pi)$.

The above corollary can be easily proved by using theorem 2 repeatedly. Furthermore, since $U_\delta f$ and $U_* f$ are monotone contraction mappings, $f^m(\pi)$ is monotone decreasing sequence. That is, by using the operator $(U_\delta f)$ iteratively, f^m converges in norm to the fixed point C^*, $i.e.$, $U_\delta C^*=C^*$. In the next section, an algorithm using $U_\delta f$ and $U_*\bar{f}$ is suggested.

4. Algorithm

In the previous section, it is shown that $U_\delta f$ is piecewise linear on Π for any simple policy whenever $f(\pi)$ is piecewise linear function on Π. Thus, the expected discounted cost of any stationary policy, δ, can be approximated by iterative use of the operator $U_\delta f$, ($i.\ e.$, successive approximation method). In the case of implementation of the operator $U_\delta f$, the cost function and policies can be specified by a finite number of items—the inequalities describing each cell of a simple partition and the corresponding action or linear function. The simple partition and the corresponding piecewise linear function are updated by the transformation function.

After predetermined iterative use of $U_\delta f$, an approximation of the expected cost of a stationary policy δ is obtained as a piecewise linear function. The concave hull of the function is used in policy improvement. The use of the concave hull of $f(\pi)$, $\bar{f}$, leads to improved policies if the approximation is sufficiently close to $f(\pi)$. Since it can be easily seen that $f'(\pi) \leq \bar{f}(\pi) \leq f(\pi)$ where $f'(\pi)$ represents the expected cost of the improved policy in policy improvement step. (See Sondik(1978)). In the reminder of this section, the algorithm is presented in general terms and the proof concerning convergence is given.

An algorithm for finding an ϵ-optimal policy starts with a simple policy δ_1 satisfying $f^1 \geq U_{\delta 1} f^1$. An iteration of the algorithm is described as follows :

At the start of the m^{th} iteration, we have a simple policy δ_m and a piecewise linear function f^m satisfying $f^m \geq U_{\delta m} f^m$, $m=0, 1. 2, \cdots$

i) Compute $U^h_{\delta m} f^m$, where the integer h is the number of iterations of $U_{\delta m}$ which are to be performed.

ii) Set $f^{m+1} = U^h_{\delta m}\, f^m$ and compute the concave hull $\bar{f}^{m+1}$ of f^{m+1}.

iii) Find a policy δ_{m+1} such that $U_{\delta m+1} \bar{f}^{m+1} = U_* \bar{f}^{m+1}$

iv) If $\|U_* \bar{f}^{m+1} - \bar{f}^{m+1}\| \leq (1-\beta)\epsilon$, then stop ϵ-optimal policy, δ^m. Otherwise, increase m by 1 and go to step i).

In the algorithm, $\|\cdot\|$ is the supremum norm. If $h=1$ in step i), the algorithm becomes the successive approximation algorithm. That is, the algorithm is modification of the policy iteration algorithm that applies the successive approximation method to the value determination step. As an example of the initial simple policy and piecewise linear function f satisfying $f \geq U_\delta f$, $\delta(\pi) = k$ for all $\pi \in \Pi$, and $f^1(\pi) = M/(1-\beta)$ for each $\pi \in \Pi$, is to be considered, where $M = \max_{i,k} q_i^k$

The following theorem shows that if the algorithm terminates then it will provide an ϵ-optimal cost function and an ϵ-optimal policy.

Theorem 3. If $\|U_* \bar{f}^m - \bar{f}^m\| \leq (1-\beta)\epsilon$, then

$$\|\bar{f}^m - C^*\| \leq \epsilon, \quad i.e., \ \bar{f}^m \text{ is } \epsilon\text{-optimal.}$$

Proof. Note that $U_* C^* = C^*$ and U_* is a contraction mapping.

$$\|\bar{f}^m - C^*\| \leq \|\bar{f}^m - U_* \bar{f}^m\| + \|U_* \bar{f}^m - U_* C^*\| \leq \|\bar{f}^m - U_* \bar{f}^m\| + \beta \|\bar{f}^m - C^*\|$$

$$(1-\beta)\|\bar{f}^m - C^*\| \leq \|\bar{f}^m - U_* \bar{f}^m\| \leq (1-\beta)\epsilon$$

Therefore, $\|\bar{f}^m - C^*\| \leq \epsilon$

That is, since the operators, $U_\delta f$ and $U_* f$, are monotone contraction mappings, the algorithm terminates with ϵ −optimal stationary policy and ϵ −optimal cost function in the finite number of iterations.

5. Example

This section gives a simple example. Consider a machine with a main component which has a probability of 0.1 that the component will break down during the manufacture of a product. For simplicity, consider only two control alternatives : $k=1$ means "examine the product" and $k=2$ means " inspect the machine". Under the action 1, we manufacture another item and examine the finished product. In the action 2, the manufacturing is interrupted for one production cycle, the machine is dismantled, and the component is inspected and replaced if it has failed. Then, the problem is infinite horizon POMDP with two states and two available alternatives. Suppose that the problem parameters are derived as Table 1. from examination cost, inspection cost, and replacement cost. The discount factor is given by $\beta=0.8$.

We arbitrarily choose the initial stationary policy as that policy minimizing the expected immediate costs of operating the process. This policy $(\delta_1)^\infty$ is simply the alternative 2 for all π and $f^1(\pi)=M/(1-\beta)=50.0$ for all π

In the first iteration, an approximation of the expected discounted cost of δ_1, $\bar{f}^2(\pi)$, constitutes of $\alpha_1=(26.35, 43.74)$, $\alpha_2=(31.27, 41.38)$, $\alpha_3=(33.48, 40.28)$. The improved policy found in the first iteration is $\delta^2(\pi)=1$ if $0 \geq \pi_1 \geq 0.754$, $\delta_2(\pi)=2$ if $0.754 < \pi_1 \leq 1$. Since $\|U_* \bar{f}^2 - \bar{f}^2\| = 3.53 > (1-\beta)\epsilon = 0.08$, the algorithm continues.

ϵ − optimal policy is found in the fifth iteration ($\|U_* \bar{f}^6 - \bar{f}^6\| = 0.064$). Thus, the ϵ − optimal policy is $\delta_5(\pi)$ obtained in the iteration 4. $\delta_5(\pi)=1$ if $0 \leq \pi_1 \leq 0.663$, $\delta_5(\pi)=2$ if not.

The expected discounted cost function of the ϵ −optimal policy consists of the following two α − vectors ; $\alpha_1=(16.37, 30.36)$, $\alpha_2=(12.75, 33.25)$.

To use this control, the controller must update π_1 after each time period using the observed states. When the action is chosen using $\delta_5(\pi)$, *i.e.*, examine the finished product if $\pi_1 \leq 0.663$ and otherwise then inspect the machine. the expected discounted cost is as shown in Figure 1, *i.e.*, $C^*(\pi) = \min_i \pi\alpha_i$

6. Conclusion

This paper has considered a stochastic optimization problem to maintain a machine with n identical internal components. The problem has formulated as infinite horizon POMDP with discount factor. This infinite model deals with discount factors which lie in the range $0 \leq \beta <$, since total expected discounted cost of all policy can be infinite if $\beta=1$.

The paper presents some useful properties of a stationary policy and finds ϵ −optimal policy by applying the successive approximation method to the value determination step of the well known policy iteration algorithm.

Table 1. Parameters for the example

alternative	P^*		q^*	B^*	
1	0.9	0.1	2.0	0.9	0.1
	0.0	1.0	5.0	0.5	0.5
2	1.0	0.0	10.0	1.0	0.0
	1.0	0.0	10.0	0.0	1.0

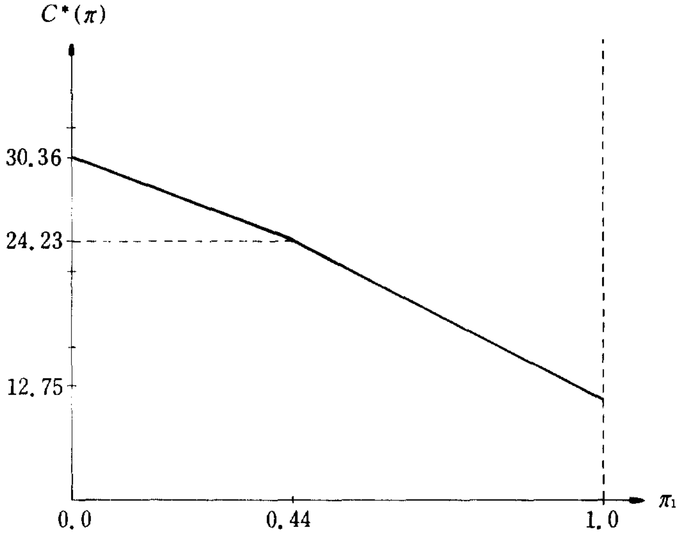

Figure 1. ε – optimal discounted expected cost

REFERENCES

1. Albright, S.C. (1979), "*Structural Results for Partially Observable Markov Decision Processes,*" Oper Res., 27, 1041~1053.
2. Platzman, L.K. (1980), "*Optimal Infinite–Horizon Undiscounted Control of Finite Probabilistic Systems,*" SIAM J.Con. & Opt. 18, 362~380.
3. Sawaki, k. (1980), Piecewise Linear Markov Decision Process with an application into Partially Observable Models, is "*Recent Developments in Markov Decision Processes,*" (R.Hartley et al, Eds.), Academic Press, New York.
4. ______ (1983), "*Transformation of Partially Observable Markov Decision Processes into Piecewise Linear Ones,*" J.Math. Anal. & Appl. 91. 112~118.
5. Tijms, H.C. and F.A. van der Duyn Schouten(1984), "*A Markov Decision Algorithm for Optimal Inspections and Revisions in a Maintenance System with Partial Information,*" Euro.J.Oper. Res., 21. 245~253.
6. Smallwood R.D. and Sondik, E.J. (1973) "*The Optimal Control of Partially Observable Markov Process over a finite Horizon,*" Oper Res., 21, 1071~1088.
7. Sondik, E.J. (1978), "*The Optimal Control of Partially Observable Markov Process over the Infinite Horizon ; Discounted Costs,*" Oper, Res., 26. 348~358, 1978.
8. Blackwell, D. (1965), "*Discounted Dynamic Programming,*" Ann. Math. Stat., 36 226~235.
9. Denardo, E.V. (1967), "*Contraction Mappings in the Theory Underlying Dynamic Programming,*" SIAM Rev. 9, 165~177.