IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 12, Issue 1 Ver. II (Jan. - Feb. 2016), PP 18-23
www.iosrjournals.org

Markov Decision Model for Maintenance Problem of Deteriorating Equipment with Policy Iteration

Hassan-Sheikh Aisha[1], DrAbubakar, U.Y[2]
[1]*Department of Mathematics and Statistics Federal Polytechnic Bida, Nigeria.*
[2]*Department of Mathematics Federal University of Technology Minna, Nigeria.*

***Abstract**: This paper analyses a dynamic system which is reviewed at equidistant points of time and at each review, the system is classified into one possible number of states and subsequently a decision has to be made. The economic consequences of the decisions taken at the review times are reflected in costs. These properties of Markov decision process are employed to study the maintenance condition of deteriorating equipment. Consequently, the optimal cost of transition from a bad condition to a good condition and the long-run fraction of time that the equipment is in bad condition were obtained. The result could be used to study the status of equipment used in various organizations to determine their efficiency and productivity.*
***Keywords**: Markov decision process, optimal cost, long-run fraction of time, decision epochs, maintenance, deterioration.*

I. Introduction

In systems where the availability of the equipment is of concern, the efficient repair and/or replacement of thisequipmentare critical to the continued usefulness of the system. This repair and replacement of equipment is called maintenance. Maintenance has a definite influence on operating costs, either through its own (maintenance) labor or through its effect on system downtime and efficiency. Maintenance can also be used to increase the probability that a system will continue to operate efficiently, given that it is allowed a certain amount of downtime for repairs. The purpose of maintenance is to return a failed or deteriorating component to a satisfactory operating state. Deterioration is a process where the condition of a component gradually worsens. If left unattended, the process will lead to deterioration failure.

In general, there are two maintenance strategies that can beapplied to return the component to a satisfactory operating state. The first strategy is to repair only when a component fails to operate or when its cost of operation becomes exorbitantly high. This is called corrective maintenance (CM), or emergency repair. The second strategy is to inspect periodically and then to repair and/or replace as is needed. This is called preventive maintenance (PM).

The purpose of PM is to eliminate the need for radical treatment sometime in the future (which is almost always much more expensive). PM, by its very nature, can be scheduled and controlled for a minimum cost. Clearly, toolittle maintenance may have very costly consequences but onthe other hand, it may not be economical to perform it too frequently. The problem of replacement or overhaul of equipment, which deteriorates with usage, is one of the standard applications of Markov processes. Continuous operation, or daily start/stop operation, without failure requires a comprehensive plant preventive maintenance planning and diagnostic system for each equipment and component. In this paper, we analyze the optimal average cost and fraction of time that equipment is in bad condition in long-run, using Markov decision process with policy iteration.

II. Markov Decision Process

We consider a dynamic system evolving over time where the probabilistic law of motion can be controlled by taking decisions. Also, costs are incurred (or rewards are earned) as a consequence of the decisions that are sequentially made when the system evolves over time. An infinite planning horizon is assumed and the goal is to find a control rule which minimizes the long-run average cost per time unit.

Consider a dynamic system which is reviewed at equidistant points of time,t= 0, 1, . . . At each review the system is classified into one of a possible number of states and subsequently a decision has to be made. The set of possible states is denoted by I . For each state $i \in I$,a set $A(i)$ of decisions or actions is given. The state space I and the action sets $A(i)$ are assumed to be *finite*. The economic consequences of the decisions taken at the review times (decision epochs) are reflected in costs. This controlled dynamic system is called a *discrete-time Markov model*when the following theMarkovianproperty is satisfied. If at a decision epoch the action,$a$is chosen in state i, then regardless of the past history of the system, the following happens:

(a) An immediate cost $c_i(a)$ is incurred,

(b) At the next decision epoch the system will be in state *j* with probability $p_{ij}(a)$, where

$$\sum_{j\in I} p_{ij}(a) = I,\ i \in I \tag{1.0}$$

Note that the one-step costs $c_i(a)$ and the one-step transition probabilities $p_{ij}(a)$ are assumed to be time homogeneous.

Assumption

For each stationary policy R, a state r (that may depend on R) exists which can be reached from any other state under policy R.

Using finiteness of the state space, as the above assumption implies that for each stationary policy R associated by Markov chain {*Xn*} satisfies the preceding two assumptions. Thus, for each stationary policy R, we have that theMarkov chain {*Xn*} has a unique equilibrium distribution $\{\pi_j(R),\ j \in I\}$. For any *j* ∈*I*, $\lim_{m\to\infty} \frac{1}{m}\sum_{n=1}^{m} p_{ij}^{(n)}(R) = \pi_j(R)$ independently of the initial state *i*. Then $\pi_j(R)$ are the unique solutions to the system of equilibrium equations $\pi_j(R) = \sum_{i\in I} p_{ij}(R_i)\pi_i(R),\quad j \in I$, (6.2.6)

(1.1)

In conjunction with $\sum_{j\in I} \pi\ j(R) = 1$ By (6.2.2), (6.2.3) and (6.2.5), (1.2)

$$g_i(R) = g(R) \text{ for all } i \in I \tag{1.3}$$

With

$$g(R) = \sum_{j\in I} cj(R)\pi_j(R) \ (6.2.7) \tag{1.4}$$

III. Policy Improvement:

A stationary policy R^* is said to be average cost optimal if $g(R^*) = g(R)$ for each stationary policy R. it has been observed that it is computationally not feasible to find an average cost optimal policy by computing the associated average cost for each stationary policy separately from equations (1.1) to (1.4). Tijms (1998)

For any stationary policy,it is assumed that the Markov chain {*Xn*} associated with policy *R* has no two disjoint closed sets. Then the average cost $g_i(R) = g(R)$, independently of the initial state *i* ∈*I.* The starting point is theobvious relation $\lim_{n\to\infty} V_n(i,R)/n = g(R)$ for all *i*, where $V_n(i,R)$ denotes thetotal expected costs over the first *n* decision epochs when the initial state is *i* andpolicy *R* is used. This relation motivates the heuristic assumption that bias values*vi(R)*, *i* ∈*I,*exist such that, for each *i* ∈*I* ,

$$V_n(i,R) \approx\ ng(R) + vi(R) \text{ for} n \text{ large.} \quad (6.2.8)\ \ (1.5)$$

Note that $vi(R) - vj(R) \approx V_n(i,R) - V_n(j,R)$ for *n* large. Thus $vi(R) - vj(R)$ measures the difference in total expected costs when starting in state *i* rather thanin state *j*, given that policy *R* is followed. This explains the name of *relative values* for the $v_i(R)$. We have the recursion equation

$$V_n(i,R) = c_i(R_i) + \sum_{j\in I} p_{ij}(R_i)V_{n-1}(j,R), \qquad n \ge 1 \quad and \ \ i \in I \tag{1.6}$$

The recursion equation follows under the condition that the next state is j, the total expected cost over the remaining next n-1 decision epoch is $V_{n-1}(j,R)$. The next state is j with probability $p_{ij}(R_i)$ when action $a = R_i$ is used in the starting state, i . Substituting the asymptotic expansion (1.5) in the recursion equation, we find, after cancelling out common terms,

$$g(R) + v_i(R) \approx c_i(R_i) + \sum_{j\in I} p_{ij}(R_i)v_j(R), \quad i \in I, \tag{1.7}$$

Which yield the value-determination equations for policy *R.*

A rigorous way of introducing the relative values associated with a given stationary policy *R* is to consider the costs incurred until the first return to some regeneration state for policy *R.* We choose some state r such that for each initial state the Markov chain {*Xn*} associated with policy *R*will visit state,*r*after finitely many transitions, regardless of the initial state. Thuswe can define, for each state $i \in I$,

$T_i(R)$ = the expected time until the first return to state *r* whenstarting in state *i* and using policy *R*.

In particular, letting a cycle be the time elapsed between two consecutive visits to the regeneration state *r* under policy *R*, we have that $T_r(R)$ is the expected length of a cycle. Also define, for each $i \in I$,

$K_i(R)$ = the expected costs incurred until the first return to state *r*when starting in state *i* and using policy *R*.

We use the convention that $K_i(R)$ includes the cost incurred when starting in state *i* but excludes the cost incurred when returning to state *r*. By the theory of renewalrewardprocesses, the average cost per time unit equals the expected costs incurred in one cycle divided by the expected length of one cycle and so

$$g(R) = \frac{K_r(R)}{T_r(R)} \qquad (1.8)$$

Next we define the particular relative value function

$$w_i(R) = K_i(R) - g(R)T_i(R), \quad i \in I \qquad (6.3.1) \qquad (1.9)$$

Note, as a consequence of (6.3.1), the normalization

$$w_i(R) = 0.$$

We state the following theorem without proof that the average cost per unit time and the relative values can be calculated simultaneously by solving system of linear equations.

Theorem 6.3.1

Let R be a given stationary policy

(a) The average cost g(R) and the relative values $w_i(R)$, $i \in I$, satisfy the followingsystem of linear equations in the unknowns g and v_i, $i \in I$,

$$v_i = c_i(R_i) - g + \sum_{j \in I} p_{ij}(R_i)v_j, \quad i \in I. \qquad (6.3.2)$$

(1.10)

(b) Let the numbers g and v_i, $i \in I$, , be any solution to (1.10). Theng = g(R)

and, for some constant c,

$$v_i = w_i(R) + c, \quad i \in I,$$

(1.11)

(c) Let s be an arbitrarily chosen state. Then the linear equations (1.10) together with the normalization equation $v_s = 0$ have a unique solution.

The economic interpretation of the relative value shows that for any solution $\{g(R), V_i(R)\}$ to the value determination equation (1.10), the numbers $v_i(R)$, $i \in I$, are called the relative values of the various starting states when policy R is used.

Assuming the Markov chain $\{X_n\}$ is aperiodic, we have for any two states $i, j \in I$,$V_i(R) - V_j(R) =$ the difference in total expected costs over an infinitely long period of time by starting in state i rather than in state j when using policy R. In other words, $V_i(R) - V_j(R)$ is the maximum amount that a person is willing to pay to start a system in state j rather than in state I when the system is controlled by rule R.

Theorem

Let g and vi , i ∈I , be given numbers. Suppose that the stationary policy R has the property

$$c_i(R_i) - g + \sum_{j \in I} p_{ij}(R_i)v_j \le v_i \quad \text{for each } i \in I, (6.2.11) \qquad (1.12)$$

Then the long-run average cost of policy R satisfies

$$g_i(R) \le g, \quad i \in I, \qquad (6.2.12) \qquad (1.13)$$

where the strict inequality sign holds in (1.13) for i = r when state r is recurrent under policy R and the strict inequality sign holds in (1.12) for i = r. The result is also true when the inequality signs in (1.12) and (1.13) are reversed.

Proof

Suppose that a control cost of $c_i(a) - g$ is incurred each time theaction,a is chosen in state i, while a terminal cost of v_j is incurred when thecontrol of the system is stopped and the system is left behind in state j. Then(1.12) states that controlling the system for one step according to rule R andstopping next is preferable to stopping directly when the initial state is i. Since thisproperty is true for each initial state, a repeated application of this property yieldsthat controlling the system for m steps according to rule R and stopping after thatis preferable to stopping directly. Thus, using the notation

$$V_m(i,R) = \sum_{t=0}^{m-1} \sum_{j\in I} p_{ij}^{t}(R) c_j(R_j) \tag{1.14}$$

(that is the expected cost to be incurred at the decision epoch t, given that $X_0 = i$ and the policy R is used). For each initial state $i \in I$,

$$V_m(i,R) - mg + \sum_{j\in I} p_{ij}^{(m)}(R) v_j \le v_i, \; m = 1,2,\ldots \tag{1.15}$$

Dividing both sides of this inequality by m and letting m→∞, we get (1.13). Next we give a formal proof that yields the result with the strict inequality sign as well. The proof is first given under the assumption that the Markov chain {Xn}associated with policy R is unichain. Then this Markov chain has a unique equilibriumdistribution {π_j(R), j ∈I}, where π_j (R) >0 only if state j is recurrent under policy R. multiply both sides of (1.12) by π_i (R) and sum over i. This gives

$$\sum_{i\in I} \pi_i(R) c_i(R_i) - g + \sum_{i\in I} \pi_i(R) \sum_{j\in I} p_{ij}(R_i) v_j(R) \le \sum_{i\in I} \pi_i(R) v_i . \tag{1.16}$$

The presents the policy iteration algorithm (Tijm 1998)

Step 1 (initialization):Choose a stationary policy R.

Step 2 (value-determination step):For the current rule R, compute the uniquesolution $\{g(R), V_i(R)\}$ to the following system of linear equations:

$$v_i = c_i(R_i) - g + \sum_{j\in I} p_{ij}(R_i) v_j \quad i \in I \tag{1.17}$$

$$v_s = 0 \tag{1.18}$$

wheres is an arbitrarily chosen state.

Step 3 (policy-improvement step):For each state i∈I , determine an action a_i yielding the minimum in

$$\min_{a\in A(i)} \left\{ c_i(a) - g(R) + \sum_{j\in I} p_{ij}(a) v_j(R) \right\} \tag{1.19}$$

The new stationary policy R is obtained by choosing $R_i = a_i$ for all i ∈I with theconvention that R_i is chosen equal to the old action R_i when this action minimizesthe policy-improvement quantity.

Step 4 (convergence test):If the new policy R = R, then the algorithm is stoppedwith policy R. Otherwise, go to step 1 with R replaced by R.

The policy-iteration algorithm converges after a finite number of iterations to anaverage cost optimal policy.

IV. The Model

At the beginning of each day a piece of equipment is inspected to reveal its actualworking condition. The equipment will be found in one of the working conditions $i = 1, 2, \ldots N$, where the working condition i is better than the working condition i + 1. The equipment deteriorates in time. If the present working condition is i and no repair is done, then at the beginning of the next day the equipment hasworking condition j with probability p_{ij} . It is assumed that $p_{ij} = 0$ for $j < i$ and $\sum_{j \geq i} p_{ij} = 1$. The working condition $i = N$ represents a malfunction that requires repair taking two days. For the intermediate states i with $1 < i < N$ there is a choice between preventively repairing the equipment and letting theequipment operate for the present day. Let the preventive repair takes only one day and a change from a bad condition to a repaired system has the working condition i = 1. We wish to determine a maintenance rule which minimizes the long-term fraction of time the equipment is in repair.

Let us put the problem in the framework of a discrete-time Markov decisionmodel. We assume a cost for each day is in repair, the long-term average cost per day represent the long-term fraction of days that the equipment is in repair. Also, since a repair for malfunction N takes two days and in the Markov decision model, the state of the systemhas to be defined at the beginning of each day, we need an auxiliary state for thesituation in which a repair is in progress. Thus theset of possible states of the system is chosen as

$$I = \{1, 2, \ldots , N , N + 1\} \quad (1.19)$$

Here State i with $1 \leq i \leq$ N corresponds to the situation in which an inspection reveals working condition i , while stateN + 1 corresponds to the situation in which a repair is in progress already for one day. Denoting the two possible actions by

$$a = \begin{cases} 1 & if\ a\ preventive\ repair\ is\ done \\ 0 & if\ no\ repair\ is\ done \end{cases} \quad (1.20)$$

The set of possible actions in state i is chosen as
A(1) = {0}, A(i)= {0, 1} for 1 < i < N, A(N) = A(N + 1) = {1}.

The one-step transition probabilities $p_{ij}(a)$ are given by

$p_{ij}(1) = 1$ for 1 < i < N

$p_{N,N+1}(1) = 1$

$p_{N+1,1}(1) = 1$

$p_{ij}(0) = p_{ij}$ for $1 \leq i < N$ and $j \geq i$,

$p_{ij}(a) = 0$ otherwise.

The one-step costs $C_i(a)$ are given by

$C_i(1) = 1$ and $C_i(0) = 0$

A policy for controlling the working condition of the equipment is a repair for taking actions at decision epoch.

Considering the Markovian assumption, and the fact that the planning horizon is infinitely long, we therefore consider stationarypolicies.

A stationary policy R is a rule that always prescribes single action$R_i$whenever the system is found in state i at decision epoch.

The rule prescribing a repair for bad equipment only when it is in good condition for at least5 working days is given by $R_i = 0$ for $1 = i \leq 5$ and $R_i = 1$ for $i = 5 = N + 1$.

V. Illustration

The average cost is optimal when the number of possible working condition of the equipment equals N = 5 and deterioration probabilities of the good equipment in a company is given below

$$P = \begin{pmatrix} 0.15 & 0.80 & 0.05 & 0 & 0 \\ 0 & 0.60 & 0.20 & 0.10 & 0.10 \\ 0 & 0 & 0.40 & 0.35 & 0.25 \\ 0 & 0 & 0 & 0.50 & 0.50 \end{pmatrix}$$

The policy iteration algorithm is initialized with the policy which prescribes repair, be it preventive repair or corrective repair, a=1 in each state except state 1 from equations (1.17) t0 (1.19), after some iterations, we obtain the minimum fraction of days that the equipment is in bad condition equals 0.214 and to have assumed a cost of one unit for each time the equipment is in repair, we therefore have that value as the average cost for optimal condition.

Conclusion

The relative value associated with the policy obtained represent both the fraction of time in the long-run that the equipment could be in bad condition and perhaps not operative, and the minimal cost incurred in repair. This could be determined for each equipment, so that the equipment whose value less compare to others could be considered as being in bad condition quite often therefore not functional and can be discarded. Note that the cost obtained is not realistic as such other methods could be used to determine the cost.

References

[1]. Abubakar, U.Y, Reju, S.A. and Awojoyogbe, B.O. (2007).Markov Decision Model and the application to the cost of treatment of Leprosy disease.Leonardo journal of sciences 11,69-78.

[2]. Abubakar, U.Y. (2011). Markov Decision Model for Human Health with Policy Iteration.Journal of science, Technology, Mathematics and Education (JOSTMED), 7(2)..

[3]. Bakker, J.D., H.J. van der Graaf, and J.M. van Noortwijk (1999).Model of Lifetime-Extending Maintenance. In M.C. Forde, editor, Proceedings of the 8thInternational Conference on Structural Faults and Repair, London, UnitedKingdom, 1999. Edinburgh: Engineering Technics Press.

[4]. Barlow, R.E., and F. Proschan (1965).Mathematical Theory of Reliability. New York: John Wiley and Sons.

[5]. Das, P.C. (1999). Prioritization of bridge maintenance needs, Case Studies in Optimal Design and Maintenance Planning of Civil Infrastructure Systems, D.M. Frangopol, editor, ASCE, Reston, Virginia, 26-44.

[6]. Derman, C. (1970).Finite State Markovian Decision Processes. New York: Academic Press

[7]. Dekker, R. (1996). Applications of maintenance optimization models: a review and analysis. Reliability Engineering and System Safety, 51:229-240.

[8]. Dekker, R., and P.A. Scarf (1998). On the impact of optimisation models in maintenance decision making: the state of the art. Reliability Engineering andSystem Safety, 60:111-119.

[9]. Frangopol, D.M., and P.C. Das (1999). Management of bridge stocks based on future reliability and maintenance costs. Bridge Design, Construction, and Maintenance, Institution of Civil Engineers, Thomas Telford, London, 45-58.

[10]. Frangopol, D.M., J.S. Kong, and E.S. Gharaibeh (2001).Reliability-based life-cycle management of highway bridges.Journal of Computing in Civil Engineering, ASCE, Vol. 15, No. 1, 27-34.

[11]. Howard, R.A. (1960), Dynamic Programming and Markov Processes. New York: Wiley.

[12]. Tijm, H.C. (1988). Stochastic Modelling and Analysis: A computational approach. New York: John Wiley & Sons.