Reliability Engineering and System Safety 154 (2016) 137–151

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Value of information in sequential decision making: Component inspection, permanent monitoring and system-level scheduling

Milad Memarzadeh, Matteo Pozzi *

Department of Civil and Environmental Engineering, Carnegie Mellon University, 107b Porter Hall, 5000 Forbes Ave., Pittsburgh, PA 15213-3890, USA

ARTICLE INFO

Article history:
Received 23 December 2015
Received in revised form
21 May 2016
Accepted 28 May 2016
Available online 6 June 2016

Keywords:
Inspection scheduling
Pre-posterior analysis
System maintenance

ABSTRACT

We illustrate how to assess the Value of Information (VoI) in sequential decision making problems modeled by Partially Observable Markov Decision Processes (POMDPs). POMDPs provide a general framework for modeling the management of infrastructure components, including operation and maintenance, when only partial or noisy observations are available; VoI is a key concept for selecting explorative actions, with application to component inspection and monitoring. Furthermore, component-level VoI can serve as an effective heuristic for assigning priorities to system-level inspection scheduling. We introduce two alternative models for the availability of information, and derive the VoI in each of those settings: the Stochastic Allocation (SA) model assumes that observations are collected with a given probability, while the Fee-based Allocation model (FA) assumes that they are available at a given cost. After presenting these models at component-level, we investigate how they perform for system-level inspection scheduling.

1. Introduction

In this paper, we investigate how to assess the Value of Information (VoI) in infrastructure management (IM), to integrate optimal control with strategies for information gathering, including component-level inspection and permanent monitoring, and system-level inspection scheduling.

We can model the IM process, including designing, maintaining, repairing and operating an infrastructure system, as a sequential decision making problem, where the manager infers and predicts the system's condition that evolves due to aging and deterioration, and takes periodic actions with the overall goal of minimizing a long-term cost function. Among the actions available to the manager, some are (at least partially) *exploratory* actions, which provide relevant information on the system state. By reducing uncertainty, information systematically improves the management process; however, as exploring is generally expensive, the manager has to find an appropriate trade-off between collecting information and improving the system state through actions. This is the so-called "exploration *vs* exploitation" problem, and its solution poses high computational challenges [1].

VoI is a utility-based metric related to decision making under uncertainty, and it measures the expected benefit due to the availability of a piece of information. In principle, all actions (both exploitative and explorative) can be treated consistently in a unified framework, without the need of directly assessing any VoI. However, it may be convenient to explicitly compute the VoI when the acquisition of a costly observation or set of observations is considered at current time. The VoI for one observation depends on all other available information since, depending on the setting, the same data can be relevant or redundant. This creates specific computational challenges in sequential decision making, when the availability of future observations usually depends on decisions that have not been taken yet. Generally, in sequential decision making the agent cannot decouple the selection of current actions to that of future ones, since the long-term cost depends on the whole policy implemented during the entire process. The output of the decision analysis, therefore, should be an optimal long-term *policy*, not just an optimal *current action*. Specifically, the optimal policy is deeply related to the available information, not only because the agent can process any collected observation and update her current belief, but also because the overall effect of a current action cannot be separated by future collection of observations and its consequences.

In this paper, we adopt the Partially Observable Markov Decision Process (POMDP) framework to model sequential decision making, because of its flexibility and generality. POMDPs include probabilistic models of degradation, cost and observation depending on the management decisions, and are rooted in Bayesian analysis. Both Markov Decision Processes (MDPs) and POMDPs have been extensively investigated for IM applications [2–13], due

* Corresponding author.
E-mail addresses: miladm@cmu.edu (M. Memarzadeh), mpozzi@cmu.edu (M. Pozzi).

http://dx.doi.org/10.1016/j.ress.2016.05.014
0951-8320/

to the computational efficiency of dynamic programming. A seminal introduction on VoI analysis is provided by the book of Raiffa and Schlaifer [14], while applications to management of structural and infrastructure systems is provided by Pozzi and Der Kiureghian [15], Straub [16], Zonta et al. [17], and Malings and Pozzi [18]; most applications refer to a single-decision making problem, however assessment of the VoI in sequential decision making is also presented in [16,19,20]. While metrics for inspection scheduling applied to IM have been developed for general applications [19,21], we recently proposed a VoI-based approximate heuristic for system-level inspection scheduling for POMDPs [22]. Also recently, Srinivasan and Parlikad [23] investigated how to evaluate VoI in POMDPs.

In this paper, we generalize our previous work, and illustrate how to integrate the VoI assessment in the POMDP framework under two specific assumptions, which we call Stochastic-based Availability (SA) and Fee-based Availability (FA), discussing in details the applicability and performance of the corresponding models. After an introduction to POMDP (Section 2) we propose a general approach to evaluate the impact of inspecting a component or permanently monitoring its condition (Section 3), and to optimize the system-level inspection scheduling under limited resources, evaluating the overall impact of inspectors (Section 4), before drawing conclusions (Section 5).

2. Sequential decision making

2.1. General setting and the MDP framework

In sequential decision-making, an agent (e.g. the infrastructure manager) selects a sequence of actions, receiving periodic rewards and possibly observations from the system she is interacting with. The scope of the process is to minimize the long-term expected cost, which is usually discounted to its present value. A classical framework for sequential decision-making is Markov Decision Process (MPD) [24], which can be efficiently solved by dynamic programming. However, an MDP assumes perfect information on the system state at any step of the decision process and, because of this, is not suitable for investigating the impact of information gathering.

2.2. The POMDP framework

The POMDP framework shares many assumptions of MDP. At any time, the system's state s assumes one value in finite discrete set $S=\{1,2,\ldots,|S|\}$, while the agent can select one action a among set $A=\{1,2,\ldots,|A|\}$. Based on the current state and action, she pays cost r (traditionally, letter r indicates a "reward", but we use it to refer to a cost). Time is discretized in steps, and variables s_t, a_t, r_t indicate state, action and cost at time t respectively. Expected cost is assigned by function $R(i,k)=\mathbb{E}[r_t|s_t=i,a_t=k]$, where $\mathbb{E}$ indicates the statistical expectation. After an action is taken, the state evolves stochastically following a Markov process governed by transition probability function $T(s,a,s')=\mathbb{P}[s_{t+1}=s'|s_t=s,a_t=a]$, where $\mathbb{P}[X]$ indicates the probability of event X.

In MDPs, action a_t follows the direct observation of the state s_t that, given the Markovian assumption, is a sufficient statistic for the process. On the contrary, POMDPs assume that at time t the agent has access only to a noisy and incomplete measure of the current state, summarized by observation z_t which can assume one value in set $Z=\{1,2,\ldots,|Z|\}$. The relation between state and observation is captured by the conditional observation probability (i.e. emission) function $O(s,a,z)=\mathbb{P}[z_t=z|s_t=s,a_{t-1}=a]$. The entire cost, transition and emission functions are listed in corresponding matrices $\mathbf{T}$, $\mathbf{O}$, $\mathbf{R}$, of size $|S|\times|S|\times|A|$, $|S|\times|Z|\times|A|$ and $|S|\times|A|$ respectively. In IM processes, the transition matrix $\mathbf{T}$ defines the degradation model and the effectiveness of maintenance actions, emission matrix $\mathbf{O}$ defines the accuracy of observations collected by instrumented and visual inspections, while cost matrix $\mathbf{R}$ defines the economic model.

Fig. 1 shows a graphical model of a POMDP, using the classical notation of dynamic Bayesian networks and influence diagrams (adopted, for example, by the textbook of Barber [25]): circles define random variables, squares decision variables, diamonds utility variables, and arrows dependence among variables. Only shaded variables are observed. Fig. 1 allows us to follow in detail the temporal process. At time t_0, hidden state is s_0, the agent takes action a_0 and pays cost r_0; then a time step passes, state evolves to s_1 that the agent observes imperfectly through z_1. Action a_1 is selected after having analysed z_1. Cost, next step state and observation depend on the action taken; and the process is iterated indefinitely.

The agent's goal is to minimize value V, defined as the expected sum of the discounted costs over an infinite time horizon: $V=\mathbb{E}\left[\sum_{t=0}^{\infty}\gamma^t r_t\right]$, using discount factor γ (it is to be noted that, in other traditional contexts, the value indicates an expected reward that has to be maximized). At time t, the agent's knowledge about the current state is represented by a probability distribution, or *belief* vector $\mathbf{b}_t$, so that its i^{th} entry is $b_t(i)=\mathbb{P}[s_t=i|\bar{z}_t,\bar{a}_t]$, with sets $\bar{a}_t=\{a_0,\ldots,a_{t-1}\}$ and $\bar{z}_t=\{z_1,\ldots,z_t\}$ being the history of observations and actions up to time t, respectively. Since the belief is a sufficient statistic for the process, the agent can base her decisions on that. Formally, a POMDP is defined by the 8-tuple $(S,Z,A,\mathbf{T},\mathbf{O},\mathbf{R},\mathbf{b}_0,\gamma)$, where $\mathbf{b}_0$ is the initial belief. In the following, we summarize its parameters in set $\boldsymbol{\Theta}=\{\mathbf{T},\mathbf{O},\mathbf{R},\gamma\}$, since the dimensions of matrices carry information on the sets S, Z and A. During the process, the agent updates her belief by iteratively processing any available observation. Transition and emission probabilities can be combined in operators that allow for predicting the state evolution and processing observations, making use of Bayes' rule. The move-forward ($\mathbf{f}$), emission ($\mathbf{e}$), and updating ($\mathbf{u}$) operators, of dimension $|S|$, $|Z|$ and $|S|$ respectively, are defined entry-by-entry as follows:

$$\begin{cases} f_i(\mathbf{b},k,\boldsymbol{\Theta})=\mathbb{P}[s_{t+1}=i|a_t=k,\mathbf{b}_t=\mathbf{b},\boldsymbol{\Theta}] &= \sum_{l=1}^{|S|} T(l,k,i)b(l) \\ e_j(\mathbf{b},k,\boldsymbol{\Theta})=\mathbb{P}[z_{t+1}=j|a_t=k,\mathbf{b}_t=\mathbf{b},\boldsymbol{\Theta}] &= \sum_{i=1}^{|S|} O(i,k,j)f_i(\mathbf{b},k,\boldsymbol{\Theta}) \\ u_i(\mathbf{b},k,j,\boldsymbol{\Theta})=\mathbb{P}[s_{t+1}=i|a_t=k,\mathbf{b}_t=\mathbf{b},\boldsymbol{\Theta},z_{t+1}=j] &= \dfrac{O(i,k,j)f_i(\mathbf{b},k,\boldsymbol{\Theta})}{e_j(\mathbf{b},k,\boldsymbol{\Theta})} \end{cases} \quad (1)$$

In summary, if the agent has belief $\mathbf{b}$ at time t, takes action k and observes j at the next step, then the updated belief is $\mathbf{u}(\mathbf{b},k,j,\boldsymbol{\Theta})$.

The agent's behaviour is defined by a *policy*, i.e. a map between belief and action. When policy π is adopted, action at time t is set as $a_t=\pi(\mathbf{b}_t)$. The value depends on policy π via the recursive linear equation:

$$V^{\pi}(\mathbf{b},\boldsymbol{\Theta})=r(\mathbf{b},\pi(\mathbf{b}),\boldsymbol{\Theta})+\gamma\sum_{z=1}^{|Z|} e_z(\mathbf{b},\pi(\mathbf{b}),\boldsymbol{\Theta})V^{\pi}[\mathbf{u}(\mathbf{b},\pi(\mathbf{b}),z,\boldsymbol{\Theta}),\boldsymbol{\Theta}] \quad (2)$$

where we re-use letter r for indicating expected immediate cost as a function of belief $\mathbf{b}$ and action a, as $r(\mathbf{b},a,\boldsymbol{\Theta})=\sum_{s=1}^{|S|} b(s)R(s,a)$. The optimal value is defined by the Bellman Equation [26]:

$$V^*(\mathbf{b},\boldsymbol{\Theta})=\min_{a\in A}\left\{r(\mathbf{b},a,\boldsymbol{\Theta})+\gamma\sum_{z=1}^{|Z|} e_z(\mathbf{b},a,\boldsymbol{\Theta})V^*[\mathbf{u}(\mathbf{b},a,z,\boldsymbol{\Theta}),\boldsymbol{\Theta}]\right\} \quad (3)$$

and it is associated with optimal policy $\pi^*(\mathbf{b},\boldsymbol{\Theta})$ that can be identified using "argmin" instead of "min" in Eq. (3).

All formulas presented in the following Sections rely on the possibility of computing value V^*, when belief $\mathbf{b}$ and parameter set