

A specialised POMDP form and algorithm for clinical patient management

Niels Peek

Medical Informatics, University of Amsterdam,
P.O. Box 22700, 1100 DE Amsterdam, The Netherlands
E-mail: n.b.peek@amc.uva.nl

Abstract

Partially observable Markov decision processes (POMDPs) have recently been suggested as a suitable model to formalising the planning of clinical patient management over a prolonged period of time. However, practical application of POMDP models is hampered by the computational complexity of associated solution methods. It is argued that the full generality of POMDPs is not needed to support many decision problems in clinical patient management, and that specialised forms are often sufficient. A specialised form of POMDP, tailored to a particular type of management problem, is introduced. It is described how a new solution method, based on Monte Carlo simulations of the decision process, can take advantage of this specialised form.

1 Introduction

Managing patients that suffer from a progressive disease is a complicated task involving a mixture of test planning, treatment selection, and prognostic assessment. The large number of possible management strategies over time precludes formalisation of this task using traditional representations such as decision trees and influence diagrams. Recently, partially observable Markov decision processes (POMDPs) [4, 8] have been suggested as a providing a suitable, integrated approach to this type of management problem [7, 10]. POMDPs are models for sequential decision making under conditions of uncertainty and limited observation opportunities. By taking into account both immediate and longterm consequences of decisions, POMDPs provide a powerful framework for decision-theoretic planning of clinical actions. Unfortunately, the computational burden associated with solving POMDPs is overwhelming, precluding their application to problems of practical size [9].

However, for many specialised problems, the full-blown generality of the POMDP approach and its associated solution methods is superfluous. We believe that this holds in particular for clinical decision problems, where often the class of admissible solutions is significantly constrained. In this paper, we discuss a specialisation of POMDPs that is tailored to a frequently re-occurring type of clinical management problem, and propose a solution method that is able to exploit the properties of this specialised form. The management problem we envision to support looks as follows. A patient suffers from a disease from which natural recovery is possible, but which may also cause harmful complications over time. There are possibilities to halt progress of the disease and its complications by intervention (e.g. surgery), but these involve a serious risk to the patient. The main problem is therefore deciding whether

or not to intervene, and if so, when. Prior to intervention, it is possible to perform several diagnostic procedures; these procedures reveal information on the clinical state of the patient at the time the procedure is undertaken, but they also comprise a (smaller) risk. A secondary problem is therefore the selection and timing of diagnostic procedures.

2 Model form

In this section, we briefly describe the general POMDP model and its associated solution form. Given a set X of variables, let Ω_X denote the set of all *configurations* of X , i.e. all possible value assignments to variables from X . A POMDP model is a tuple (T, X, A, P, o, L) , where

- T is a linearly ordered set of *decision moments*,
- X is a finite set of *stochastic variables*, jointly defining the set Ω_X of *states*,
- A is a finite set of available *actions*,
- $P = \{p_t^a : \Omega_X \times \Omega_X \rightarrow [0, 1] \mid a \in A, t \in T\}$ is a set of time- and action-dependent *transition probability functions*,
- $o : A \rightarrow \wp(X)$ is an *observation function*, and
- $L : \{l_t : \Omega_X \times A \rightarrow \mathbb{R} \mid t \in T\}$ is a set of time-dependent *loss functions*.

The set T of decision moments denotes the points in time where the decision maker is expected to select an action $a \in A$. We restrict ourselves to *finite-horizon problems*, and take $T = \{0, 1, 2, \dots, N\} \subset \mathbb{N}$. No action is selected at the last decision moment $t = N$; this moment is included for evaluation of the final state only. The clinical state of the patient is described by the set X of discrete, stochastic variables; let $\mathcal{S} = \Omega_X \times \dots \times \Omega_X = \Omega_X^{N+1}$ denote the set of all possible state sequences. When configuration $c \in \Omega_X$ characterises the state at time point $t \in T$, selection of action $a \in A$ will result in a transition to state c' at time point $t + 1$ with probability $p_t^a(c, c')$. Furthermore, the decision maker is able to observe the configuration of the set $o(a) \subseteq X$ at time point t , and can use this observation to optimise subsequent decision making; note that the observation function o is independent of time. At each decision moment the decision maker also incurs a loss $l_t(c, a)$; the losses associated with subsequent moments in a realisation of the decision process are combined by a *utility function* $u : \mathbb{R}^{N+1} \rightarrow \mathbb{R}$.

Now, let ϕ be a joint probability distribution on X at the initial time point $t = 0$, reflecting the decision maker's prior beliefs on the clinical state of the patient. Given ϕ and a sequence α of action choices for all decision moments (except $t = N$), we obtain a probability distribution on $\Pr_{\phi, \alpha}$ on the set \mathcal{S} of possible state sequences. From this distribution, we can compute the expected utility of action sequence α under ϕ . Our objective is to select actions during the decision process such that expected utility is maximised. Prior to the first action choice, we therefore compose a *decision-theoretic plan* π , which prescribes an action choice for each time point $t < N$, given the history of past actions and observations. When m is the maximum number of distinct observations that may follow an action choice (i.e., if $Y = o(a)$ then $|\Omega_Y| \leq m$, for each action $a \in A$), we have that m^N is an upper bound on the size decision-theoretic plans. The number of possible plans is bounded by k^{mN} , where

$k = |A|$ is the number of available actions. It is therefore not surprising that the problem of finding the optimal plan is PSPACE-complete [9].

A POMDP model was recently developed to support the clinical management of patients with *ventricular septal defect* (VSD), a frequently occurring congenital heart disease [10]. A VSD is an abnormal opening in the heart causing heart failure and associated symptoms such as shortness of breath, feeding problems, and growth retardation. Approximately 70% of all VSDs close spontaneously in the first years of life due to tissue growth, obviating the need for surgical intervention. However, in the long run the disease may cause irreversible damage to the lungs and a severely impaired respiratory function. Several diagnostic tests (ECG, echocardiography, cardiac catheterisation, chest X-ray, and pulmonary biopsy) are available to examine the patient’s condition before deciding upon cardiac surgery. The cardiologist treating a VSD patient therefore faces the type of management problem described in the previous section. The POMDP model for VSD has 33 state variables, yielding approximately $9,7 \cdot 10^{15}$ possible configurations (i.e. states of the POMDP); many configurations, however, cannot occur in practice, or can only occur in specific circumstances. This is expressed in the transition probability functions by assigning zero probability to those configurations. The model distinguishes 6 decision moments (ages of the patient, ranging from 3 months to 8 years), and 7 distinct actions to choose from.

It was also shown in [10] how *temporal probabilistic networks* can be used to graphically represent the transition probability functions of a POMDP model, and how this representation, by exploiting conditional independence relations between state variables, can strongly reduce the number of probability estimates required to complete the model. This is especially useful when the number of state variables is large: the complexity of transition probability functions quickly grows in the number of variables. A compact representation, as in probabilistic networks, is then indispensable as obtaining probability estimates is often a cumbersome task. We do not further elaborate on this representation here, and refer the interested reader to the paper in question for more details.

3 A specialised POMDP form

We will now propose a special form of POMDP model that is tailored to support the management problem described in Section 1. We first characterise the types of loss and utility function that are used within this special form, and then describe three restricting assumptions we make on actions, transition probabilities, and plan structure.

We take a loss $l_t(c, a)$, $t < N$, to represent the *mortality risk* associated with state c and action a at time point t , and a loss $l_N(c')$ to denote *life expectancy* (in years) associated with final state c' at time point $t = N$, where no action choice is made. Let r_0, \dots, r_{N-1} be such mortality risks, obtained from a given evolution of the decision process (i.e. states and actions for each of the decision moments up to time point $t = N - 1$). Then,

$$s_t = \prod_{i=0}^{t-1} (1 - r_i) \tag{1}$$

denotes the chance that the patient survives at least up to time point $t > 0$. Now, let d_t be the (fixed) actual duration (in years) between the start of the decision process and decision

moment t , $0 \leq t \leq N$; we then have that

$$le_t = \sum_{j=1}^{t-1} d_j r_j s_j \quad (2)$$

is the life expectancy of the patient up to time point t . The following utility function u now expresses overall life expectancy:

$$u(r_0, \dots, r_N) = le_N + (d_N + r_N) \cdot s_N, \quad (3)$$

where $r_N = l_N(c')$ denotes life expectancy at the final time point. This type of utility function is generally referred to as *risk-sensitive* [5]. We note that it is also possible to encode mortality risks in the transition probability functions, but we deliberately choose not to do so, for reasons explained shortly.

We make three further assumptions on the POMDP model and its admissible solutions. First, the set A is taken to be composed of three disjoint sets A_{test} , A_{treat} , and A_{skip} , where A_{test} constitutes the set of available diagnostic procedures, A_{treat} lists treatment alternatives, and A_{skip} is a singleton set that consists of the special action *skip* (i.e. refrain from acting at the specified point in time) only. The set A_{treat} is assumed to be relatively small compared to A_{test} ; e.g., in the VSD domain, we have $A_{\text{treat}} = \{\textit{surgery}\}$ and $A_{\text{test}} = \{\textit{ECG}, \textit{echo}, \textit{catheter}, \textit{X-ray}, \textit{biopsy}\}$. Second, from A_{treat} an action is selected at most once, and after that moment, further action is refrained from (by selecting *skip* for all subsequent moments). Before the moment of treatment though, actions may be selected freely from A_{test} and A_{skip} . From the first two assumptions we thus obtain a restricted set Π of admissible plans, in each of which there is but a single moment of control, preceded by multiple moments of observation. The size of the set Π is bounded by $(k_{\text{test}} + 1)^{mN}$, where $k_{\text{test}} = |A_{\text{test}}|$, and as before, m is the maximum number of distinct observations that may follow an action choice. Although $k_{\text{test}} < k$ (where $k = |A|$), this number of admissible plans is still very large. The *average* size of plans in Π , however, equals $m^{N/2}$.

The third and last assumption is that state development is independent of test actions. So, $p_t^a = p_t^{\textit{skip}}$ for each $a \in A_{\text{test}}$, $t = 0, \dots, N - 1$. Note that we can make this assumption because mortality risks are encoded in the loss functions: this enables us let all diagnostic procedures induce the same transition probabilities, even if they differ with respect to their associated risks. Without this assumption, each of the k^N possible action sequences α induces a different probability distribution $\Pr_{\phi, \alpha}$ on state sequences. With the assumption, many action sequences induce the same distribution: we obtain $(k_{\text{treat}} + 1)^N$ classes of action sequences, $k_{\text{treat}} = |A_{\text{treat}}|$, where the sequences in each class induce the same distribution. Action sequences that are obtained from one of the admissible plans in the set Π though, contain at most one action choice from A_{treat} . With that restriction, the number of classes therefore further reduces to $N \cdot k_{\text{treat}} + 1$. We will exploit this significant reduction in the solution method described below.

4 Solution method

The standard approach to solving POMDP problems was initiated by Aström [1] and Sondik [11], and is based on transforming the POMDP into an equivalent, fully observable Markov decision process (called the *belief MDP*), over all possible probability distributions on the

original state space Ω_X . The belief MDP can be solved using value iteration, a form of dynamic programming [2]. However, the continuous state space of the belief MDP is computationally difficult to handle, and therefore the associated solution algorithms are complicated and limited [8]. Notwithstanding recent algorithmic advances in this field [3, 6], solving POMDP problems of considerable size with this approach seems to be infeasible; the current state of the art allows to solve POMDPs with at most 10 to 15 states. Another disadvantage of dynamic programming is that the decisions are optimised in reverse order. This implies that we cannot exploit prior knowledge of the problem involved (e.g. patient-specific information), and it is difficult to take into account constraints on plan structure, as for instance occur in the specialised POMDP form described above. We therefore propose a new solution method to solve POMDPs, tailored to the specialised form described above. Due to space limitations, we restrict ourselves to giving a sketch of the proposed method.

Basically, our method estimates expectations of the utility function u under a given decision-theoretic plan $\pi \in \Pi$ by simulating the stochastic process on X under plan π . These Monte Carlo estimates are then compared to establish the optimal plan. With this approach, we can easily exploit prior knowledge of the problem case, as each simulation starts from the initial decision moment; this is especially useful when many potential state sequences are ruled out by the initial state. Constraints on plan structure are taken into account by selecting plans from the admissible set Π only. Furthermore, we can take advantage of the fact that the distribution on state sequences is fixed by the choice and timing of treatment. Let $\sigma_1, \dots, \sigma_n$ be independent and identically distributed samples from \mathcal{S} , where treatment action $a \in A_{\text{treat}}$ was selected at time point $t < N$ in the simulations. Since the transition probabilities are equal for all test actions and the skip action, we can use these samples to estimate expectations of the function u for all action sequences that select treatment a at moment t , regardless of their prior testing policy. So, the simulation effort is strongly reduced as we evaluate a large variety of action sequences from a single collection of samples.

The space Π of admissible plans will generally be too large to enumerate. We therefore perform a local search through Π , stepwise refining the plan under consideration. The search process proceeds as follows. Let ϕ represent given beliefs on the initial state, and let α be the action sequence where action $a \in A_{\text{treat}}$ is selected at time point $t < N$, and *skip* is selected at all other times. Note that α also represents a (rather unsophisticated) plan $\pi \in \Pi$: ‘perform action a at time point t without prior testing’. Now, let $\Pr_{\phi, \alpha}$ as before be the distribution on \mathcal{S} induced by ϕ and α , and let S be a collection of independent and identically distributed samples from \mathcal{S} drawn using $\Pr_{\phi, \alpha}$. If $\hat{u}(\sigma, \alpha)$ denotes the life expectancy associated with state and action sequences σ and α , then

$$\bar{u}_{\phi, \alpha}(S) = \frac{1}{|S|} \sum_{\sigma \in S} \hat{u}(\sigma, \alpha) \quad (4)$$

is an Monte Carlo estimate of life expectancy under plan π . To obtain more sophisticated plans, we now try to find *indicators* of variation in \bar{u} . We say that the set $Y \subseteq X$ is such an indicator at time point $t' < t$, if there exists configurations c'_Y and c''_Y of Y such that difference between $\bar{u}_{\phi, \alpha}(S')$ and $\bar{u}_{\phi, \alpha}(S'')$ is statistically significant, where S', S'' are the subcollections of state sequences matching c'_Y and c''_Y at time point t' , respectively. We restrict the search process to indicators Y that are *observable*, i.e. $Y = o(a')$ for some action $a' \in A_{\text{test}}$. Furthermore, the difference between estimated life expectancies must remain significant when adjusted for performing test action a' at time point t' . The plan π is now refined by adding

the test action corresponding to the indicator that induces the most significant difference in life-expectancy estimates. Subsequently, the treatment action and its timing are re-considered for each of the possible observations that may follow a' ; new simulations may be needed to obtain the necessary samples here. After possible adjustment of the treatment choice under each of the observations, the process is repeated; policy refinement is halted when no further improvements can be found.

We note that Monte Carlo estimates converge to correct expected values in the limit of taking an infinite number of samples. In practice, however, a finite, and often small, number of samples is sufficient. Furthermore, the number of samples corresponding to particular events is balanced with the likelihood of these events to occur. In our application of the technique, this means that highly improbable state developments are considered only after taking a large number of samples. At the start of the policy-refinement process, improvements to the policy will be based on developments that are either very likely to occur or induce large differences in life expectancy. As the refinement process proceeds and more samples are obtained, improvements may also be based on rare developments that induce small differences.

5 Discussion and future work

POMDPs provide a powerful modelling framework for decision-theoretic planning, with promising applications to multi-stage clinical decision problems. The generality of the standard POMDP model, however, limits practical application of the framework due to the computational complexity of associated solution methods. To alleviate this obstacle, we have proposed a specialised POMDP form and algorithm to support a frequently encountered type of clinical management problem. The specialised form assumes several restrictions on the effects of actions on state development, and on the structure of admissible solutions. These restrictions jointly reduce the number of action-sequence classes that induce a different probability distribution on state sequences. Our algorithm exploits this property by reducing the simulation effort in Monte-Carlo evaluation of decision-theoretic plans: each sample of the stochastic process is used to evaluate a large number of action sequences.

We are currently implementing our algorithm, and plan to evaluate its performance on the VSD model in the near future. Further research is required to investigate extensions to the basic model form proposed here. For instance, more elaborate loss and utility functions that incorporate quality of life and costs of treatment, are needed to provide a more realistic account of the tradeoffs in real-world clinical decisions. Furthermore, allowing a larger number of control moments is needed to support a wider range of management problems. To prevent a combinatorial explosion in the solution space, this extension should be coped to a fine-grained classification of action types and associated restrictions on admissible treatment plans.

Acknowledgements

The investigations were (partly) supported by the Netherlands Computer Science Research Foundation with financial support from the Netherlands Organisation for Scientific Research (NWO). The author wishes to thank Jaap Ottenkamp for his support in constructing the VSD model.

References

- [1] K.J. Aström. Optimal control of Markov processes with incomplete state information. *J. Math. Anal. Appl.*, 10:174–205, 1965.
- [2] R.E. Bellman. *Dynamic Programming*. Princeton University Press, 1957.
- [3] A.R. Cassandra, M.L. Littman, and N.L. Zhang. Incremental pruning: a simple, fast, exact method for partially observable Markov decision processes. In *Proc. 13th Conf. Uncertainty in Artificial Intelligence (UAI-97)*, pp. 54–61, 1997.
- [4] A.W. Drake. *Observation of a Markov Process through a Noisy Channel*. Ph.D. thesis, MIT, 1962.
- [5] E. Fernández-Gaucherand and S.I. Marcus. Risk-sensitive optimal control of hidden Markov models: Structural results. *IEEE Trans. Automatic Control*, 42:1418–1422, 1997.
- [6] M. Hauskrecht. Incremental methods for computing bounds in partially observable Markov decision processes. In *Proc. 14th Nat. Conf. Artif. Intell. (AAAI-97)*, 1997.
- [7] M. Hauskrecht and H. Fraser. Planning medical therapy using partially observable Markov decision processes. In *Proc. 9th Int. WS Principles of Diagnosis (DX-98)*, pp. 182–189, 1998.
- [8] W.S. Lovejoy. A survey of algorithmic methods for partially observed Markov decision processes. *Ann. Oper. Res.*, 28:47–66, 1991.
- [9] C.H. Papadimitriou and J.N. Tsitsiklis. The complexity of Markov decision processes. *Math. Oper. Res.*, 12(3):441–450, 1987.
- [10] N.B. Peek. Explicit temporal models for decision-theoretic planning of clinical management. *Artif. Intell. Med.*, 15(2):135–154, 1999.
- [11] E.J. Sondik. *The Optimal Control of Partially Observable Markov Processes*. Ph.D. thesis, Stanford University, 1971.