Modeling Treatment of Ischemic Heart Disease with Partially Observable Markov Decision Processes.

Milos Hauskrecht[1], Hamish Fraser[2,3]

[1] Computer Science Department, Box 1910, Brown University, Providence, RI 02912 *milos@cs.brown.edu*
[2] Tufts-New England Medical Center, 750 Washington Street, Boston, MA 02111
[3] MIT Laboratory for Computer Science, 545 Technology Square, Cambridge, MA 02139 *hamish@medg.lcs.mit.edu*

Abstract

Diagnosis of a disease and its treatment are not separate, one-shot activities. Instead they are very often dependent and interleaved over time, mostly due to uncertainty about the underlying disease, uncertainty associated with the response of a patient to the treatment and varying cost of different diagnostic (investigative) and treatment procedures. The framework of Partially observable Markov decision processes (POMDPs) developed and used in operations research, control theory and artificial intelligence communities is particularly suitable for modeling such a complex decision process. In the paper, we show how the POMDP framework could be used to model and solve the problem of the management of patients with ischemic heart disease, and point out modeling advantages of the framework over standard decision formalisms.

INTRODUCTION

The diagnosis of a disease and its treatment are not separate processes. Although the correct diagnosis helps to narrow the appropropriate treatment choices it is often the case that the treatment must be pursued without knowing the underlying patient state with certainty. The reason for this is that the diagnostic process is not a one shot activity and it usually neccessary to collect additional information about the underlying disease, which in turn may delay the treatment and make the patients' outcome worse. This process is often even more complex, when uncertainty associated with the reaction of a patient to different treatment choices and costs associated with various diagnostic (investigative) actions need to be considered. Thus in a course of patient management one needs to carefully evaluate the benefit of possible diagnostic (investigative) and treatment steps and their ordering with regard to the overall global objective, the well being of a patient. The management of patients with ischemic heart disease (IHD) (see e.g. [1]) is an example of such a problem and we will focus on it in our work.

To model accurately the complex sequential decision process that combines diagnostic and treatment steps we need a framework that is expressive enough to capture all relevant features of the problem. The tools typically used to model and analyze decision processes are (stochastic) decision trees (see e.g. [2]), Markov decision processes (MDPs) [3; 4] or semi-MDPs [5]. Although all of these allow us to model the uncertainty associated with the outcome of the treatment, they fail to capture processes in which a state describing the underlying disease is hidden (unknown) and is observed only indirectly via a collection of incomplete or imperfect observations. This feature is crucial for many therapy problems, in which an underlying disease cannot be identified with certainty and thus more options need to be considered.

A framework more suitable for modeling the outlined therapy problem is Partially observable Markov decision process (POMDP) [6] [7]. POMDP represents a controlled Markov process, similar to Markov decision process [3; 4], and it explicitly represents two sources of uncertainty: stochasticity related to the dynamics of the control process (outcome of the treatment or diagnostic procedure is not deterministic), and uncertainty associated with the partial observability of the disease process by a decision-maker (the underlying disease state is observed indirectly via incomplete or imperfect observations). The objectives of the control (treatment) are modeled by means of a reward or cost model, that represents payoffs associated with different situations, and temporal objective function that combines rewards obtained over multiple steps. A POMDP model is usually easier to build and modify (compared to the decision tree) and it can be used also for other tasks, such as prediction and explanation.

PARTIALLY OBSERVABLE MARKOV DECISION PROCESS

A *partially observable Markov decision process (POMDP)* describes a stochastic control process with partially observable states and formally corresponds to a 6-tuple (S, A, Θ, T, O, R), where S is a finite set of process states (disease states); A is a finite set of actions (diagnostic and treatment procedures); Θ is a finite set of observations (findings, results of diagnostic tests); $T : S \times A \times S \rightarrow [0, 1]$ is a set of transition probabilities between states that describe the dynamics of the modeled system; $O : S \times A \times \Theta \rightarrow [0, 1]$ stand for a set of observation probabilities that describe the relationship among observations, states and actions; and $R : S \times A \times S \rightarrow \mathcal{R}$ denotes a reward (cost) model that assigns rewards to state transitions and models payoffs associated with such transitions.

Given a POMDP model, the objective is to construct a *policy* that prescribes how a decision-maker should act in order to maximize expected cumulative reward over some horizon of interest. Two types of decision models are typical: *finite horizon* where we adopt a policy that maximizes $E(\sum_{t=0}^{T} r_t)$, such that r_t is a reward obtained at time t, and *infinite horizon, discounted* problem, where a policy that maximizes $E(\sum_{t=0}^{\infty} \gamma^t r_t)$, with $0 < \gamma < 1$ being a discount factor, is sought. We will focus our attention on the infinite horizon discounted model.

In POMDPs process states are hidden and decisions could be based only on observations seen and past actions performed. This makes a large difference when the optimal policy for all possible situations a decision-maker may encounter should be found. While in perfectly observable Markov processes [3], [4] one works with a finite number of states that are always known, in POMDPs underlying states are not known with certainty and one has to work with and base all decisions on belief states. A belief state assigns a probability to every possible state $s \in S$, and there is an infinite number of possible belief states one may encounter. Value (expected discounted reward) of the optimal policy for a belief state b satisfies the Bellman's equation [3]:

$$V^*(b) = \max_{a \in A} R(b,a) + \gamma \sum_{o \in \Theta} P(s|b,a) V^*(\tau(b,a,o)), \tag{1}$$

where b stands for the belief state, $R(b,a)$ denotes an expected one step reward for a belief state b and an action a and equals:

$$R(b,a) = \sum_{s' \in S} \sum_{s \in S} R(s,a,s') P(s'|s,a) b(s),$$

and τ is an update function that computes a new belief state given a previous step belief, action and observation:

$$\tau(b,a,o)(s) = \beta P(o|s,a) \sum_{s' \in S} P(s|s',a) b(s'),$$

with β being a normalizing constant. The optimal action for a belief state is then obtained as:

$$\mu^*(b) = \arg\max_{a \in A} R(b,a) + \gamma \sum_{o \in \Theta} P(s|b,a) V^*(\tau(b,a,o))$$

In general equation 1 for V^* could be divided into two parts: an expected immediate reward for performing action a in a belief state b — $R(b,a)$ and an expected reward for following the optimal policy afterwards. V^* can be approximated using value iteration strategy [3]. In this strategy, we define an i-th step approximation as:

$$V_i(b) = \max_{a \in A} R(b,a) + \gamma \sum_{o \in \Theta} P(s|b,a) V_{i-1}(\tau(b,a,o)).$$

The sequence of value function approximations is guaranteed to converge to the optimal solution. Note that unfolding of the formula for some initial b could be represented also by a decision tree [8]. In this case, belief states and not observations must be used and expected rewards from multiple steps should be combined and considered.

The important property of the approximation sequence is that value functions V_i are piecewise linear and convex [7], which allows us to compute the update in finite time for the complete belief space. Although computable, the computational cost for doing so is high and only smaller belief-state MDPs could be solved exactly or to arbitrary small precision. In practice, this leads to various approximation methods that allow us to compute good solutions fast [8; 9].

APPLYING POMDPS TO MEDICAL THERAPY PLANNING

The expressivness of the POMDP framework makes it suitable for therapy planning problems with hidden disease states and with both diagnostic and treatment steps. An example of such a problem, with complex temporal dependencies, is the problem of management of patients with ischemic heart disease (IHD) [1].

Management of ischemic heart disease

Ischemic heart disease is caused by an imbalance between the supply and demand of oxygen to the heart. The condition is most often caused by the narrowing of coronary arteries (coronary artery disease) and an associated reduction in the oxygenated blood flow. The coronary artery disease tends to progress over time. The pace of the disease progress is stochastic and contigent on multiple factors.

At any point in time the physician has different options to intervene: do nothing, treat the patient with medication, perform surgical procedures (angioplasty — PTCA, coronary artery bypass surgery — CABG), or one of the investigative procedures (angiogram, stress test) that tend to reveal more about the underlying status of the coronary disease. Some of the interventions have a low cost, but some carry a significant cost associated with the invasiveness of the procedure.

The objective of the therapy planning is to develop a strategy that would minimize the expected cummulative cost of the treatment, where the cost is defined in terms of the dead-alive tradeoff, quality of life, invasiveness of procedures and their economic cost. The optimal strategy depends not only on the immediate action choice, but also on future choices, thus reflecting complex temporal tradeoffs.

POMDP for IHD

Although standard POMDP formalism matches well the characteristics and needs of many therapy planning prob-

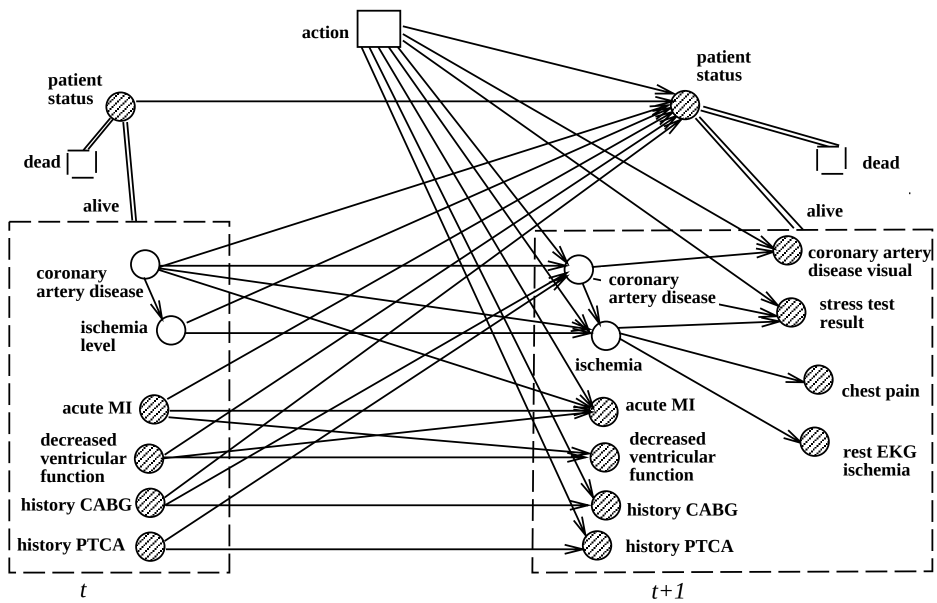

Figure 1: Model of the dynamics for the ischemic heart disease. State variables are represented by circles, an action choice as a rectangle. Patterned circles are used for observables. For time t, only process state variables are shown.

lems, it is often hard to apply it directly to these problems, due to the size and complexity of models one has to face. This complicates the task of model definition (parameter acquisition) and it is often also beyond current limits of exact POMDP problem-solving methods. To overcome these problems for the IHD domain we built a more compact version of POMDP and focused on various approximate problem-solving methods.

Model of dynamics

To model the dynamics of the IHD disease process more compactly we used a hierarchical refinement of the dynamic belief network [10] in figure 1. In this model a state of the patient is described using a set of state variables and their associated values, e.g. moderate *coronary artery disease*, severe *chest pain*, positive *rest EKG result*, etc. State variables can be either hidden or observable, e.g., variables representing status of the *coronary artery disease*, and *ischemia level* are hidden (not observable directly), while other state variables like *chest pain*, *rest EKG result*, or *stress test result* are perfectly observable. [1] Some of the observations are unconditional and assumed to be available at any point in time (e.g. *chest pain*), while others are conditioned on an appropriate action (e.g. *stress test result* is available only when stress test procedure was chosen and performed in the previous step). A novel feature of our IHD model is that its state variable set is not flat, but hierarchically structured. More specifically, state variable *patient status*, with two values (dead or alive) enables (activates) state variables providing more detailed description of the patient state (as e.g. *chest pain*) only when patient is known to be alive.

In a belief network model transitions between two consecutive states can be described more compactly by taking advantage of the independence structure, which reduces the number of parameters one has to define. The parameters (probabilities) of our current model are based mostly on the model published in [1], the remaining parameters were extrapolated from the available results or estimated using clinical experience. [2]

Cost model

The cost model we use for the IHD domain is represented more compactly as well. The model consists of two components:

$$R(s, a, s') = R(s') + R(a),$$

where $R(s')$ is a cost associated with a patient state only and $R(a)$ stands for a cost associated with an action (e.g. cost of performing coronary bypass surgery that includes the economic cost, patient's discomfort, and so forth). The cost associated with a patient state, $R(s')$, can be further broken down into state variable costs. That is, a cost for a patient state can be decomposed to costs associated with an

[1] Chest pain can be classified as none, mild-moderate, and severe and it is assumed to be always observable by a physician.

[2] Note, that methods for learning probabilities [11] could be applied to acquire or refine estimates directly from data.

amount of chest pain the patient suffers at given time, occurence of myocardial infarction, etc. In such a case $R(s')$ can be expressed as:

$$R(s') = \sum_i R(s'_i).$$

where s'_i represents a value assigned to variable i. The above decomposition of the cost model reduced the number of parameters we had to estimate, which greatly simplified the model building process. The cost values were defined subjectively and represent a combined measure of the dead-alive trade-off, quality of life and economic cost.

Decision problem

The goal in the IHD therapy problem is to find the sequence of actions that minimizes expected discounted cost over infinite horizon: $\min E(\sum_{t=0}^{\infty} \gamma^t r_t)$. [3] This allows us to express longer term goals and not restrict the decision horizon to a finite number of steps. An interesting feature of our model is that we use transitions of different durations for different actions — transitions associated with surgical and investigative procedures occur within a day, and transitions associated with non-invasive actions (*no-action* and *medication*) are assumed in 3 month periods. To account for this difference, we use discounting ($\gamma = 0.95$) for long-term actions (*no-action* and *medication*); all other short-term actions are undiscounted and their costs are added fully within the model.

SOLVING THE IHD PROBLEM

The POMDP, once defined, could be converted into the *belief-state MDP*, in which beliefs over all possible state variable value combinations are assumed. Then Bellman's equation 1 holds and either exact [7; 8] or approximate value iteration techniques [9; 8] developed for the standard POMDPs could be applied. However, we can also take advantage of the additional structure present in the problem. We considered three improvements that help to reduce the complexity of problem-solving procedures for IHD and similar medical problems [8].

The first improvement stems from the fact that not all state variables are neccessary to define the belief state. For example, in the IHD problem, it is sufficient to use a belief state defined only over state variables that directly mediate transitions: *patient status*, *coronary artery disease, ischemia level*, *acute MI*, *decreased ventricular function, history of CABG* and *history of PTCA*. We call these variables *process (or information) state variables*. The second improvement is that when some of the process variables are observable, they should be treated that way and their exact values instead of beliefs over their values should be used. In the IHD problem state variables (*acute MI, decreased ventricular function, history of CABG* and *history of PTCA*) are perfectly observable, thus the state is better modeled as a hybrid state with two components: a vector of observable process variable values; and a belief over all possible combinations of values of hidden process variables. The third improvement, takes advantage of the hierarchically structured set of variables, which restricts certain state variable combinations. For example, when the patient is dead, values of other state variables are not relevant and belief over possible combinations of their values should not be considered.

Solutions

We have implemented and tested a set value function approximation methods with additional structural improvements. Table 1 illustrates a sequence of recommendations for a single patient case with a follow-up obtained for two of the best-performing methods from [8]. The first method is the incremental linear vector method with 15 incremental linear vector update cycles. Using this method, the value function for all possible belief states was computed off-line in about 30 minutes on SPARC-10 in Lucid Common Lisp. The second method tested is the fast informed bound method and it took about 3 minutes. For every stage, the table shows a set of current observations and a list of possible actions, ordered with regard to the obtained cost score. The top (lowest cost) action is executed at each step. Note that both methods always suggest the same action choices.

EVALUATION

POMDP problems of large complexity and their solutions are often hard to evaluate. The main reason for this is that we usually do not have access to optimal solutions. Thus results were compared to expert opinions of good management strategies.

The IHD model we built is of moderate complexity and includes many simplifications. Interestingly, despite that and the need to estimate a large number of parameters, the model and obtained solutions demonstrated behavior that was in most instances clinically reasonable and justifiable. This is very promising for the future work and further refinement of the model. The evaluation, and tests on patient cases also revealed current model deficiencies that require further improvements. However, these appear to be caused by the model simplification, and failure to represent all relevant details of the patient state. For example, we have found that the current model should include a variable representing a physical condition of a patient, which influences the likelihood of reaching a non-diagnostic result for the stress test procedure. Omitting this variable lead in some instances to a repeated choice of the stress-test procedure when the patient failed the procedure in the previous step.

[3] Note that costs can be be represented as negative rewards and formulas for maximizations apply.

step	current patient status	actions	score (method 1)	score (method 2)
0	chest pain: mild-moderate;	**stress-test**	**285.22**	**248.53**
	rest EKG ischemia: negative; acute MI: false;	no action	285.62	249.82
	decreased ventricular function: false;	medication	286.75	250.98
	coronary artery visual: not available;	PTCA	288.75	252.36
	stres test result: not available;	angiogram	292.92	256.68
	history CABG: false; history PTCA: false	CABG	491.94	427.77
1	chest pain: mild-moderate;	**PTCA**	**298.47**	**262.54**
	rest EKG ischemia: negative; acute MI: false	stress test	316.39	280.33
	decreased ventricular function: false;	no action	321.92	288.24
	coronary artery visual: not-available;	medication	322.72	289.12
	stress test result: positive;	angiogram	323.79	287.91
	history CABG: false; history PTCA: false	CABG	503.73	440.77
2	chest pain: no chest pain;	**no action**	**259.07**	**226.23**
	rest EKG ischemia: negative; acute MI: false;	medication	260.62	227.78
	decreased ventricular function: false;	stress test	264.35	229.87
	coronary artery visual: normal;	angiogram	273.34	239.16
	stress test result: not available;	PTCA	276.98	243.24
	history CABG: false; history PTCA: true	CABG	481.36	417.28
3	chest pain: mild-moderate;	**medication**	**451.50**	**418.07**
	rest EKG ischemia: negative; acute MI: true	no action	452.81	419.47
	decreased ventricular function: false	PTCA	464.58	429.87
	coronary artery visual: not available	angiogram	470.62	435.62
	stress test result: not available	stress-test	479.68	445.22
	history CABG: false; history PTCA: true	CABG	657.77	608.11

Table 1: Patient case with a follow-up from [8]. Recommendations for each state (in bold) are based on the lowest value function approximation (cost score) computed by the incremental linear vector method (method 1) and the fast informed bound method (method 2). Note that both methods suggest the same action choices.

CONCLUSION

The partially observable Markov decision process provides a framework suitable for modeling medical therapy planning problems and overcomes some of the modeling deficiencies of standard decision techniques. To investigate this we applied the framework to the problem of management of patients with ischemic heart disease (IHD).

This problem requires the consideration of the hidden disease state, both investigations and management strategies, and the cost and benefits of actions over multiple time stages. The solutions obtained for the IHD therapy planning domain are promising and showed that POMDP could provide a useful framework for modeling and analyzing the complex decision process. This justifies further refinement and extension of the current IHD model as well as the application of the framework to other complex decision problems.

Acknowledgements

This research was supported by the grant 1T15LM07092 from the National Library of Medicine. Authors would like to thank Peter Szolovits and William Long for their valuable feedback.

References

[1] J.B. Wong, F.A. Sonnenberg, D.N. Salem, S.G. Pauker. Myocardial revascularization for chronic stable angina. Annals of Internal Medicine, 113 (1), pp. 852-871, 1990.

[2] D.K. Owens, H.C. Sox. Medical Decision Making: Probabilistic Medical Reasoning. In E.H. Shortliffe, L.E. Perreault (eds.), Medical Informatics: Computer Applications in Health Care, Addison Wesley, 1990.

[3] E.R. Bellman. Dynamic Programming. Princeton University Press, 1957.

[4] M. Puterman. Markov Decision Processes. John Wiley and Sons, 1994.

[5] T.-Y. Leong. Dynamic decision making in medicine: a critique of existing formalisms. In Proceedings of the 17-th Annual Symposium on Computer Applications in Medical Care. pp. 478-484, 1993.

[6] K.J. Astrom. Optimal control of Markov decision processes with incomplete state estimation. Journal of Mathematical Analysis and Applications, 10, pp. 174-205, 1965.

[7] R.D. Smallwood, E.J. Sondik. The optimal control of partially observable processes over a finite horizon. Operations Research, 21, pp. 1071-1088, 1973.

[8] M. Hauskrecht. Planning and control in stochastic domains with imperfect information. MIT-LCS-TR-738, 1997.

[9] W.S. Lovejoy. A survey of algorithmic methods for partially observed Markov decision processes. Annals of Operations Research, 28, pp. 47-66, 1991.

[10] U. Kjaerulff. A computational scheme for reasoning in dynamic probabilistic networks. In the Proceedings of UAI-92, pp. 121-129, 1992.

[11] D. Heckerman. A Tutorial on Learning with Bayesian Networks. MSR-TR-95-06, Microsoft Research, 1996.