Planning Medical Therapy Using Partially Observable Markov Decision Processes.

Milos Hauskrecht
Computer Science Department, Box 1910
Brown University
Providence, RI 02912-1210
milos@cs.brown.edu

Hamish Fraser
Tufts-New England Medical Center
750 Washington Street
Boston, MA 02111
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 treatment and diagnostic (investigative) procedures. The framework particularly suitable for modeling such a complex therapy decision process is Partially observable Markov decision process (POMDP). Unfortunately the problem of finding the optimal therapy within the standard POMDP framework is also computationally very costly. In this paper we investigate various structural extensions of the standard POMDP framework and approximation methods which allow us to simplify model construction process for larger therapy problems and to solve them faster. A therapy problem we target specifically is the management of patients with ischemic heart disease.

1 Introduction

The diagnosis of a disease and its treatment are not separate processes. Although the correct diagnosis helps to narrow the appropriate treatment choices it is often the case that the treatment must be pursued without knowing the underlying patient state with certainty. The main reason for this is that the diagnostic process is not a one shot activity and it usually requires to collect additional information about the underlying disease, which in turn may delay the treatment and make the patients' outcome worse. Further complexity stems from uncertainty associated with the reaction of a patient to different treatment choices and costs associated with various diagnostic (investigative) actions. Thus in a course of patient management one needs to carefully evaluate the benefit of diagnostic (investigative) and treatment steps with regard to the overall global objective, the well being of a patient. An example of such a problem is the management of patients with ischemic heart disease (IHD) [13], and we will focus on it in our work.

To model accurately the complex decision process that combines diagnostic and treatment steps we need a framework that is expressive enough to capture all relevant features of the problem. A framework particularly suitable for such a problem is Partially observable Markov decision process (POMDP) [1] [11]. POMDP represents a controlled process, similar to MDP [2] [10], but with two sources of uncertainty: uncertainty related to the dynamics of the control process (same as MDPs), and uncertainty associated with the partial observability of the process by a control agent. The objectives of the control (treatment) are modeled by means of a reward (cost) model, that represents payoffs associated with different situations, and objective function that combines rewards obtained over multiple steps. The goal is then to find the control that optimizes the objective function.

Although the standard POMDP framework allows us to capture the main characteristics of the medical therapy problems, it can be only rarely applied and used directly. This is because: (1) the complexity of the standard model (number of parameters one needs to define) can be enormous (2) exact methods for solving the problem are computationally very demanding and only very small problems can be solved exactly. We address these problems in context of the treatment of the IHD.

To deal with the issue of model complexity we described the POMDP for the IHD problem in more compact form, using a hierarchical version of a dynamic belief network to represent the dynamics of the disease and factored hierarchically structured cost model to represent payoffs associated with different states and actions. This allowed us to take advantage of various structural regularities and independencies present in the problem domain, and to subsequently reduce the number of parameters needed to define the model. To reduce the computational complexity of problem-solving methods we applied: (1) a number of structure-based refinements, namely the ability to reduce the number of state variables we need to consider while planning (so called information state variables), the ability to separate perfectly and partially observable state components, and an additional model of state constraints; (2) approximation methods, that trade off the optimal for an approximate solution. Using structured approaches and approximations we were able to construct an IHD model of moderate complexity and solve it, which is promising for further application of the POMDP to this and other medical therapy problems.

In the following, we will first describe the basics of the POMDP framework and computational problems associated with solving problems defined in it. Then we will focus on the application domain — the management of patients with ischemic heart disease, and describe the model we built to represent the problem. After that we will discuss issues of problem solving and show how problem-solving procedures for POMDPs could be adapted to handle additional structure.

2 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 states; A is a finite set of actions; Θ is a finite set of observations; $T : S \times A \times S \to [0, 1]$ is a set of transition probabilities between states that describe the dynamic behavior of the modeled environment; $O : S \times A \times \Theta \to [0, 1]$ stand for a set of observation probabilities that describe the relationship among observations, states and actions; and $R : S \times A \times S \to \mathcal{R}$ denotes a reward (cost) model that assigns rewards to state transitions and models payoffs associated with such transitions. [1]

Given a POMDP, the objective is to construct a *policy* that maximizes expected cumulative reward over some horizon of interest. Two types of decision criteria are typical: *finite horizon* where we adopt a policy that maximizes $E(\sum_{t=0}^{\infty} r^t)$, 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. In the following we will focus our attention only on the infinite discounted horizon problem.

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 MDPs (see [2], [10]) one works with a finite number of states that are always known, and the optimal policy can be found efficiently using dynamic programming techniques, in POMDPs underlying states are not known with certainty and POMDP must be first transformed into an information-state MDP, most often into a belief-state MDP. A belief state assigns a probability to every possible state $s \in S$ and there is an infinite number of belief states one may encounter. The infinite number of possible belief states is the main problem complicating the optimization task.

Value (expected discounted reward) of the optimal policy for a belief state satisfies the Bellman's equation [2]:

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

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

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

The optimal value function in 1 could be approximated for all possible belief states using value iteration strategy [2]. In this strategy, we define an i-th step approximation as:

$$V_i(b) = \max_{a \in A} R(b, a) + \gamma \sum_{o \in \Theta} P(s|b, a) V_{i-1}(\tau(b, a, o)), \quad (2)$$

starting from some initial value function V_0.

The sequence of value function approximations converges to the optimal (fixed point) solution. The important property of the sequence is that value functions are piecewise linear and convex [11]. This makes it possible to compute the update in finite time for the complete belief space. Although computable, the computational cost for doing so is enormous and only small belief-state MDPs could be solved exactly or to arbitrary small precision in practice. This naturally leads to various approximation methods that try to calculate good solution fast (see [9], [6]). Thus the difference between perfectly observable and partially observable MDPs reflects the tradeoff between expressivity of the framework and computational complexity of the associated problem-solving procedures.

3 Modeling medical therapy planning using POMDPs

The POMDP framework is suitable for many therapy planning problems, particularly those with hidden disease states and both diagnostic and treatment steps. An example of such a problem, with complex temporal stochastic dependencies, and costs associated with both treatment and investigative procedures, is the problem of management of patients with ischemic heart disease (IHD) (see e.g. [13]).

[1] Note that costs can be treated as negative rewards.

3.1 Management of the ischemic heart disease

Ischemic heart disease is caused by an imbalance between the supply and demand of oxygen by a heart muscle. 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 choose 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, including pain and the risk of death.

The objective of the therapy planning is to come up with a strategy that would minimize the expected integral cost of the treatment, where the cost is defined in terms of the death of the patient, quality of life (chest pain) over a period of time, invasiveness of surgical and investigative procedures and their economic cost. Note that the optimal strategy depends stongly not only on the immediate action choice, but also on future choices, thus reflecting complex temporal tradeoffs.

3.2 Applying POMDPs to IHD

POMDP formalism matches well the characteristics and needs of the IHD therapy planning problem. However, it is hard to use standard POMDPs for the problem directly, because of the size and complexity of the model, which complicates the task of model definition (parameter acquisition) and is also beyond current limits of exact POMDP problem-solving methods. To overcome these problems we used a more compact version of POMDPs and focused on various approximate problem-solving methods. We start with the issue of modeling the structure.

3.3 Model of dynamics

To model the dynamics of the IHD disease process more compactly we used a hierarchical refinement of the dynamic belief network (see [5], [8]) in figure 1. In this model a state of the patient is described using a set of state variables, which can be either hidden or observable. For example, variables representing status of the *coronary artery disease*, and *ischemia level* are hidden, while other state variables like *chest pain*, *rest EKG result*, or *stress test result* are perfectly observable. Some of the observations are unconditional and assumed to be available at any point in time (e.g. *chest pain*), while others are triggered (enabled) only with an appropriate action (e.g. *stress test result* is available only when stress test procedure is chosen and performed).

The advantage of a belief network model of the system dynamics is twofold: by describing the state using a set of state variables the state at any point in time could have both perfectly and partially observable components; transitions between consecutive states can be described more compactly by taking advantage of the independence structure, which effectively reduces the number of parameters one has to define. An interesting and novel feature of the IHD model we developed 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 only when patient is known to be alive. Clearly, evaluating values for such variables does not make sense when patient is dead and when their values do not have any influence on the decision process.

The hierarchical variable structure serves two purposes. First, the fact that some of the state variables are enabled or disabled based on values of other variables effectively constrain the state space one has to consider. Second, the variable structure can be used to represent new sets of independencies within the dynamic belief network, which in turn simplifies its definition (by means of the number of parameters one has to provide). For example, in our model, the transition probability between a specific previous state (prs) and a state in which the patient is alive and suffers from a severe coronary artery disease (CAD) is represented using two probability distributions:

$$\begin{aligned}&P(status = alive, CAD = severe|action, prs) =\\ &\qquad P(status = alive|action, prs)\\ &\qquad P(CAD = severe|action, prs, status = alive).\end{aligned}$$

The first probability distribution concerns the variable *status* and represents the distribution of a patient being alive as a result of some procedure performed in the previous state. This distribution depends on values of state variables in the previous time step. The second distribution represents a conditional distribution of *coronary artery disease* given a previous state, an action and patient being alive. Such a distribution can exploit a different set of independences, i.e. the value of the *coronary artery disease* variable being independent of some previous state variables when the patient is alive. In general, our hierarchical belief network allows us to decompose the probability distribution and represent different structural independences on different levels.

3.4 Cost model

The cost (reward) model for the IHD domain can be represented more compactly as well. Our structured cost model consists of two components:

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

where $R(s')$ is a cost associated with the resulting state, e.g. a cost associated with a live patient who is suffering from

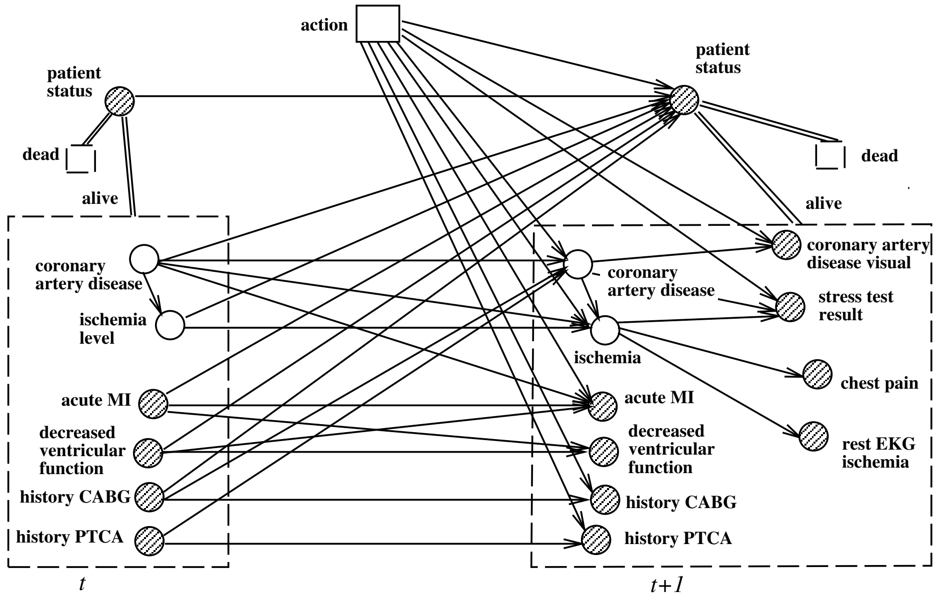

Figure 1: Model of the dynamics for the ischemic heart disease. State variables are represented by cicles, an action choice as a rectangle. Patterned circles are used for observables. For time t, only process state variables are shown.

severe chest pain and has experienced an acute myocardial infarction (acute MI); and where $R(a)$ stands for a cost associated with an action choice, e.g. a cost of performing coronary bypass surgery that includes the economic cost, patient's discomfort, and so forth. Note that two structured models (dynamics and costs) can be represented together using dynamic influence diagrams framework [12].

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 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 decomposition along state variables must respect hierarchical constraints (see above) and different costs can be associated with state variables at different levels.

The fact that the cost model decomposes into atomic costs associated with state variables and actions significantly reduces the number of parameters we need to define. This in turn simplifies the stage of building a POMDP model in which quantitative cost estimates need to be found and assesed. Once the estimates are collected, the expected one step cost associated with action a and state configuration s is:

$$R(s,a) = R(a) + \sum_{s'} p(s'|s,a)R(s')$$

3.5 Decision problem

The goal in the IHD case is to find a policy that optimizes the chosen objective function. For that we use the infinite discounted horizon model that minimizes expected discounted cost over infinite horizon. 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*) occur within 3 months. To account for this difference, we use discounting ($\gamma = 0.95$) only for long-term actions (*no-action* and *medication*); all other short-term actions are undiscounted and their costs are added fully within the model.

4 Solving IHD problem

The factored POMDP model, once defined, could be converted into the *information-state MDP*, in which beliefs over all possible state variable value combinations are assumed. Then fixed point equation 1 holds and either exact [11] [4] or approximate value iteration techniques [6] developed for the standard POMDPs could be applied. However, we can also take advantage of the additional structure present in our IHD model. In the following we will present three structural improvements based on [6]: the ability to reduce the number of state variables we need to consider during planning, the ability to separate perfectly and partially observable state components, and an additional model of state constraints.

4.1 Process state variables

The first improvement stems from the fact that not all variables representing the state of the patient at a given time are neccessary to define the information 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*. In general, the state can be represented using a local belief network with many variables. Process state variables represent their smaller subset which is consistent with the Bellman's equation and preserves the Markov property of the information process.

4.2 Hybrid information state

The second improvement takes advantage of the fact that process state variables can be both hidden and observable. In particular, 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. For example, in the IHD problem four of the process state variables (*acute MI, decreased ventricular function, history of CABG* and *history of PTCA*) are perfectly observable. In such a case the information state is better modeled as a hybrid state $\{o_d, b_d\}$ with two components: a vector of observable process variable values (o_d); and a belief over all possible combinations of values of hidden process variables (b_d).

4.3 Value function decomposition

It is well known that for the standard belief space MDP the value function one would obtain after a finite number of value iteration updates is piecewise linear and convex [11]. However, when a hybrid information state space is used, the value function can be represented as a collection of piecewise linear and convex value functions of smaller complexity, one function for each combination of perfectly observable process state variable values [6].

Let o be a vector of all observations, $proj_d(o)$ be a projection of the vector o to the vector of process state observable values, and $\{h_1, h_2, \cdots, h_j, \cdots, h_q\}$ be a set of all possible combinations of values of hidden process state variables. The value function for i-th step approximation can be rewritten as [7]:

$$V_i(\{o_d, b_d\}) = V_i^{o_d}(b_d).$$

$V_i^{o_d}(b_d)$ denotes a partial value function for a vector o_d and equals:

$$V_i^{o_d}(b_d) = \max_{\alpha \in \Gamma_i^{o_d}} \sum_{j=1}^{q} \alpha(h_j) b_d(h_j),$$

where $\Gamma_i^{o_d}$ is a set of linear vectors defining $V_i^{o_d}$. The update formula for an observation vector o_d and a belief b_d then is:

$$\begin{aligned} V_i^{o_d}(b_d) &= \max_{a \in A} \sum_{j=1}^{q} b_d(h_j)[R(o_d, h_j, a) \\ &\quad + \gamma \sum_{o' \in \Theta} \sum_{k=1}^{q} P(h_k, o'|h_j, a) \alpha_{i-1}^{o_d, b_d, o', a}], \end{aligned}$$

where $\alpha_{i-1}^{o_d,b_d,o',a}$ is a linear vector from $\Gamma_{i-1}^{proj_d(o')}$ such that:

$$\begin{aligned} \alpha_{i-1}^{o_d,b_d,o',a} &= \\ &\arg\max_{\alpha_{i-1} \in \Gamma_{i-1}^{proj_d(o')}} \sum_{j=1}^{q} \sum_{k=1}^{q} P(h_k, o'|h_j, a) b_d(h_j) \alpha_{i-1}(h_k). \end{aligned}$$

The fact that for each combination of observables the (partial) value function is piecewise linear and convex can be used to adapt both exact and approximation algorithms developed for the standard POMDP case also to hybrid circumstances. The main benefit is that the complexity of value functions is smaller, which in turn influences the complexity of problem-solving procedures. In contrast to this if we use a POMDP model with a flat state space, each (composite) state combines both observable and hidden components and we have to work with beliefs over all possible composite states. We note that other structural refinements of a value function description are possible as well. For example, Boutilier and Poole [3] proposed and explored a method that uses compact representations of linear vectors defining piecewise linear and convex value function.

4.4 Exploiting additional state constraints

One drawback of the factored state representation is that it does not provide means for restricting combinations of state variable values that are not possible. [2] Thus, if we know that some combinations of values are not possible and we want to take that into an account we need an additional model of constraints on state variables. In the IHD case we model these constraints using a hierarchically structured set of variables which explicitly restricts certain state variable combinations. For example, when a patient is dead, values of other state variables are not relevant and a belief over possible combinations of their values is not considered. [3]

The hierarchically structured set of state variables, representing constraints on possible combinations of state variable values can be nicely adapted into our hybrid framework. For example, in the IHD problem, the set of all possible combinations of observable process variable values and the number of their associated hidden variable value combinations

[2] Note that in a POMDP with flat states any redundand state can be directly excluded from the state space and not considered.

[3] The situation in which a state variable should be disabled can be always modeled using a distinct value *disabled*. However, in a case of hierarchy we can avoid modeling this value explicitly.

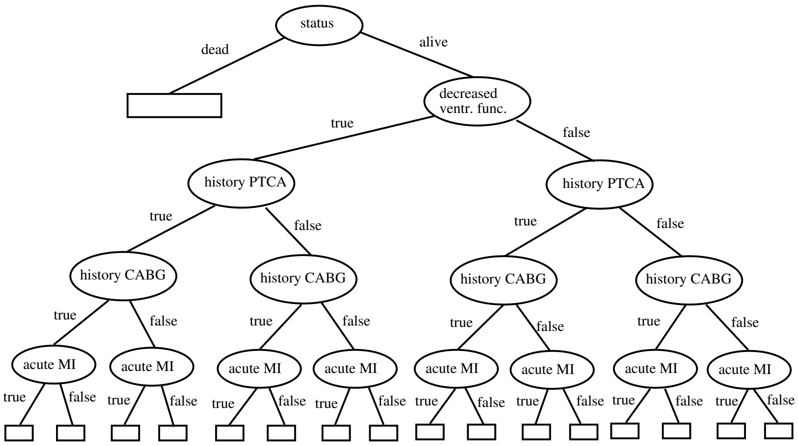

Figure 2: An information state space for the IHD problem represented using the decision tree. Internal nodes correspond to process state observables and leaf nodes to beliefs over all possible assignments to hidden process state variables consistent with observables. The structure of the tree is sparse and reflects constraints imposed by the hierarchical variable set.

may differ depending on the value assigned to variable *status*. That is, when *status* is dead all other state variables are disabled and a hybrid information state corresponds to *status=dead* only (no other hidden or observable variables are used). On the other hand when *status* is alive other state variables become relevant and the hybrid state consists of a vector of all observable variable values and a belief over all possible hidden variable combinations. The information state space for the IHD is therefore sparse and could be represented more compactly, e.g. using the decision tree structure in figure 2. [4] In general the hierarchical constraints can lead to situations in which hybrid information states use different observable and belief components.

4.5 Solutions

We have implemented and tested a set of 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 [6]. 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 was computed in 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.

4.6 Evaluation

POMDP problems of large complexity and their solutions are often hard to evaluate. The main reason is that we usually do not have an access to optimal solutions. Thus one is able to evaluate different solutions (acquired by different approximations methods) only via simulation and comparisons of simulation results. Therapy problems are often even harder, as treatment objectives, especially their combinations are hard to model and they are often subjective. Despite that, there seems to be a pretty good sense of what constitutes a clinically reasonable and justifiable choice in many instances. Thus a model and solutions obtained for it could be tested and compared against expert opinion. Such an evaluation helps to establish if the modeling framework is suitable, practically useful and worth further investigation.

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 the behavior that was in most instances clinically reasonable and justifiable. This is very promising for the future work and further extension and refinement of the model. The evaluation, and tests on multiple patient cases also revealed current model deficiencies that require further improvements. However, these are mostly caused by the model simplification, and failure to represent all relevant features of the state. For example, we have found that the current model should include also a variable representing a physical condition of a patient, which influences the likelihood of reaching a non-diagnostic result

[4] Decision trees were used also in [3] to represent linear vectors of piecewise linear and convex value functions more compactly. In fact, the ideas of separation of perfectly observable components, exploitation of additional state constraints and compact representations of linear vectors could be combined to refine the value function description. We leave this for future work.

step	current patient status		actions	cost score method 1	cost score method 2
0	chest pain:	mild-moderate	**stress-test**	285.22	248.53
	rest EKG ischemia:	negative	no action	285.62	249.82
	decreased ventr. func.:	false	medication	286.75	250.98
	acute MI:	false	PTCA	288.75	252.36
	coronary artery visual:	not available	angiogram	292.92	256.68
	stress test results:	not available	CABG	491.94	427.77
	history CABG:	false			
	history PTCA:	false			
1	chest pain:	mild-moderate	**PTCA**	298.47	262.54
	rest EKG ischemia:	negative	stress test	316.39	280.33
	decreased ventr. func.:	false	no action	321.92	288.24
	acute MI:	false	medication	322.72	289.12
	coronary artery visual:	not-available	angiogram	323.79	287.91
	stress test results:	positive	CABG	503.73	440.77
	history CABG:	false			
	history PTCA:	false			
2	chest pain:	no chest pain	**no action**	259.07	226.23
	rest EKG ischemia:	negative	medication	260.62	227.78
	decreased ventr. func.:	false	stress test	264.35	229.87
	acute MI:	false	angiogram	273.34	239.16
	coronary artery visual:	normal	PTCA	276.98	243.24
	stress test results:	not available	CABG	481.36	417.28
	history CABG:	false			
	history PTCA:	true			
3	chest pain:	mild-moderate	**medication**	451.50	418.07
	rest EKG ischemia:	negative	no action	452.81	419.47
	decreased ventr. func.:	false	PTCA	464.58	429.87
	acute MI:	true	angiogram	470.62	435.62
	coronary artery visual:	not available	stress-test	479.68	445.22
	stress test results:	not available	CABG	657.77	608.11
	history CABG:	false			
	history PTCA:	true			
4	chest pain:	mild-moderate	**PTCA**	471.16	433.98
	rest EKG ischemia:	negative	medication	483.11	447.85
	decreased ventr. func.:	true	no action	485.15	450.04
	acute MI:	false	stress-test	486.32	448.75
	coronary artery visual:	not available	angiogram	496.38	458.87
	stress test results:	not available	CABG	661.98	610.81
	history CABG:	false			
	history PTCA:	true			

Table 1: Patient case with a follow-up from [6]. Recommendations for each state (in bold) are based on the value function approximation 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.

for the stress test procedure. In the current model, the chance of a stress-test failure is treated as a random variable with higher probablity being assigned (based on population study) to one of the diagnostic outcomes and not conditioned on the physical state of the patient. This lead in some instances to a repeated choice of the stress-test procedure when the patient failed the procedure, simply reflecting the fact that it is worthwhile to flip the coin with a larger probability of successful outcome again.

5 Conclusion

Partially observable Markov decision processes provide a theoretical framework suitable for modeling many medical therapy planning problems. However, the complexity of such problems and complexity of exact problem-solving procedures for solving POMDPs often require us to seek techniques and methods that would make the framework also practically useful. In our work, we focused on two possible solutions to this problem: representation and exploitation of additional problem structure and approximation of the optimal solutions. We explored them in context of the problem of management of patients with ischemic heart disease.

The POMDP model we built for the IHD problem uses a factored state space, a hierarchical version of a Bayesian belief network to represent the system dynamics and more compact hierarchically structured cost model. Such a model allowed us to take advantage of various regularities and independencies present in the problem domain and to simplify the model construction. Additional model structure can be used not only to simplify the stage of model construction, but also to improve the performance of problem-solving procedures. In our we work we showed how to simplify the problem-solving by considering only a relevant subset of state variables, by differentiating observable and hidden variables, and by using additional hierarchical state variable contraints. Other structural approaches (see e.g. [3]) or combinations of structural approaches are possible. However, structural improvements are not always sufficient to adequately speed-up the problem-solving process. In such cases approximation methods that trade-off the solution quality for speed need to be employed. To do this for the IHD we applied various approximation techniques developed in [6].

Using structured approach and approximations we were able to built an IHD model of moderate complexity and to solve it. The solutions obtained for the current model are promising and confirmed that POMDPs could provide a useful framework for the IHD domain. Thus further extension of the model is warrantied. However, a more complex model would likely require further exploration of new structural refinements and new approximation techniques. An example could be a combination of two decision models: a one-step decision model, with higher granularity and detailedness, evaluating immediate consequences of different action choices; and simpler, more coarse POMDP model that would capture the temporal aspect of the planning process, thus providing a good estimate of expected values for states considered in the first model.

References

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

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

[3] C. Boutlier, D. Poole. Computing Optimal Policies for Partially Observable Decision Processes. In the Proceedings of AAAI-96, pp. 1168-1175, 1996.

[4] A. R. Cassandra. Exact and Approximate Algorithms for Partially Observable Markov Decision Processes. Ph. D. Thesis. Brown University, Providence, RI, 1997.

[5] T. Dean, K. Kanazawa. The model for reasoning about persistence and causation. Computational Intelligence, 5(3), pp. 142-150, 1989.

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

[7] M. Hauskrecht. Combining perfectly and partially observable Markov decision processes. under preparation, 1998.

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

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

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

[12] J.A. Tatman, R.D. Schachter. Dynamic programming and influence diagrams. IEEE Transactions on Systems, Man and Cybernetics, 20(2), pp. 365-379, 1990.

[13] J.B. Wong et.al. Myocardial revascularization for chronic stable angina. Annals of Internal Medicine, 113 (1), pp. 852-871, 1990.