A Survey of POMDP Applications

Anthony R. Cassandra
arc@mcc.com
Microelectronics and Computer Technology Corporation
Austin, TX

Abstract

An increasing number of researchers in many areas are becoming interested in the application of the partially observable Markov decision process (POMDP) model to problems with hidden state. This model can account for both state transition and observation uncertainty. The majority of recent research interest in the POMDP model has been in the artificial intelligence community and as such, has been applied in a limited range of domains. The main purpose of this paper is show the wider applicability of the model by way of surveying the potential application areas for POMDPs.

Introduction

Markov decision process models (MDPs) have proven to be useful in a variety of sequential planning applications where it is crucial to account for uncertainty in the process (Puterman 1994). The success of the application of MDP can be attributed to the existence of efficient algorithms for finding optimal solutions for MDP models (Puterman 1994).

The partially observable MDP model (POMDP) generalizes the MDP model to allow for even more forms of uncertainty to be accounted for in the process. However, the computational complexity of algorithms for optimally solving these models has mostly confined the interest in these models to the research journals (Papadimitriou & Tsitsiklis 1987).

Recently, the research community has shown increasing interest in looking for more efficient algorithms for solving POMDPs, which has naturally led to exploring various forms of approximate solution techniques. Since approximations by nature involve trade-offs, guiding applications are crucial for determining how these trade-offs should be made. Making trade-offs simply for their mathematical or computational convenience does not always translate into algorithms that are useful.

Aside from the need for applications to guide algorithm development, the recent advances in POMDP algorithms has pushed the state-of-the-art to the point where they may be or could soon be usefully deployed in some real-world settings. These improvements range from simple heuristics (**?**) to approximate forms of dynamic programming (**?**) to tighter upper and lower bounds for search heuristics (**?**).

This paper surveys some of the application areas where POMDP models can be applied. They range from smaller problems where the current technology may be sufficient to large-scale problems that are well beyond the range of the existing algorithms. Our grouping of the application areas is somewhat arbitrary due to the overlapping nature of the problems that need to be solved in the various domains.

POMDP Model

Before discussing the specific application areas, we briefly present the basic form of the POMDP model. The POMDP model we consider consists of the following:

- a finite set of states, $\mathcal{S}$;
- a finite set of actions, $\mathcal{A}$;
- a finite set of observations, $\mathcal{Z}$;
- a state transition function, $\tau : \mathcal{S} \times \mathcal{A} \rightarrow \Pi(\mathcal{S})$, where $\Pi(\cdot)$ is a probability distribution over some finite set;
- an observation function $o : \mathcal{S} \times \mathcal{A} \rightarrow \Pi(\mathcal{Z})$;
- an immediate reward function $r : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$.

The state set represents all the possible underlying states the process can be in, though this state is not directly observable. The action set is all the available control choices at each point in time. The observation set consists of all the possible observations that the process can emit. The state transition function encodes the uncertainty in the process state evolution, while the observation function relates the process outputs (observations) to the true underlying state of the process. Finally, the reward function gives the immediate utility for performing an action in each of the underlying process states.

We want to use this model to derive a control policy that will yield the greatest amount of utility over some number of decision steps. Since there is not space to include the details of the model, theory and algorithms we refer the reader to other works (Smallwood & Sondik 1973; Lovejoy 1991; White 1991; Cassandra, Kaelbling,

& Littman 1994). This coarse definition will suffice to allow us to explain how the various domains presented can be modeled as POMDPs.

Industrial Applications

Machine Maintenance

We begin our discussion of applications with domain of machine maintenance. A *machine* in this context could be any piece of mechanical equipment which requires periodic maintenance due to deterioration of its internal components over time. Although the machine can be dismantled and each part inspected to determine the full internal state of the machine, this requires an expenditure of time and personnel while rendering the machine unproductive for the duration of the inspection. For this application, we would like to obtain an inspection/replacement policy that either minimizes the operating costs or maximizes the production capacity of the machine.

The states for a POMDP model is the internal state of the components of the machine. This could be on a component-by-component basis, or some more abstract description of the machine's internal state. the actions of the model could correspond to performing maintenance, replacing components, replacing the machine, continue operating the machine or inspection. There could be a number of differnt inspection types which could vary both in their cost and their effectiveness.

The observations of the POMDP model could incorporate both the performance of the machine and the various outcomes that each inspection action would yield. The POMDP model is particularly applicable to these problems because the production and inspection observations are only probabilistically related to the internal states. For example, there are many reasons why a manufacturing machine could produce a faulty product even though its internal parts are functioning properly. Also, the least expensive inspection techniques are usually the least thorough.

In the general case, a control policy would depend upon the observations received at each point in time. Some of the earliest work using the POMDP model concerned themselves with machine maintenance problems (Eckles 1968; Pollock 1970; Ross 1971; Smallwood & Sondik 1973; Pierskalla & Voelker 1976; Rosenfeld 1976; White 1977; 1979).

Structural Inspection

Closely related to the machine maintenance application are structural inspection applications. These would include inspection and maintenance of paved roads, bridges, buildings, aircraft parts, etc (Ellis, Jiang, & Corotis 1995). Over time, the materials themselves or the critical joining components can deteriorate.

Similar to the machine maintenance application, the states would correspond to the internal composition of the parts/materials of interest. Here actions would consist of various inspection options as well as replacement or reinforcement choices.

Unlike the machine maintenance examples, structural components are not easily dismantled. This causes less than optimal inspection conditions and contributes to the uncertainty in the inspection results. Effective control policies in these domains are particularly important, since structural failures here can involve the loss of life, rather than merely a reduction in revenue.

Elevator Control Policies

Although not life-critical, good control policies for systems of elevators is another application area for POMDPs (Crites 1996). The states of the model, in this case, would be the position and direction of the elevators and the number and location of passengers waiting to be serviced. The actions consist of the various choices for where to send the individual elevators, which floors they should stop at, and which floors should they pass.

Typical elevator systems will have two directional buttons at each floor and floor-call buttons inside each elevator. The selections of these buttons become the observations available to make decisions. Because these do not provide enough information to determine the number of passengers, or their exact destinations, the system is partially observable. We could reduce the amount of partial observability by adding more buttons, extra sensing hardware, etc. to provide more specific information. However, this incurs additional costs and although it reduces the observation uncertainty, it does not eliminate it. The only way to get a completely observable system would be to mandate that every passenger indicate their exact destination; hardly a realistic or friendly system.

Fishery Industry

The population of various types of marine life in a given area is never known with certainty. Observations are made which are probabilistically related to the true population (Lane 1989). Controlling the population of fish is important in the fishery industry, so that a proper balance of near-term and long-term objectives are reached. Actions available can include restocking, imposing fishing bans, changing fishing limits, etc. With a POMDP model of the population growth, the effects of the various actions on the population, and the relative accuracies of their sampling methods, a model can be developed and a policy can be computed to help guide the decisions of the agencies responsible for maintaining this delicate balance.

Scientific Applications

Autonomous Robots

Autonomous and semi-autonomous robots are useful in a variety of hazardous or difficult to access environments, including:

- interplanetary rovers,

- deep-space navigation,
- bomb disposal,
- land-mine clearing,
- toxic waste clean-up,
- radioactive material handling,
- deep-ocean exploration,
- sewage/drainage network inspection and repair,
- etc.

For many of these, tele-operated solutions are currently used. However, if communication is difficult or impossible, or if the time-lag on the communication response is significant, tele-operated solutions are cumbersome and time-consuming. The more autonomy the robot has, the more it can accomplish and the less burden there is on the operators. Even where tele-operated solutions are feasible, added autonomy can make the robot more productive, easier to operate and less prone to human error.

No matter the quality or quantity of sensing hardware deployed on the robot, from the robot's point of view, it will have a horrendously incomplete view of its surroundings. With this ubiquitous partial observability, the POMDP model can provide the formal basis for autonomous behavior in these domains, where:

- the states are the robots location, surroundings and internal state,
- the actions are the available actuators,
- the observations are the outputs of the sensing equipment, and
- the immediate reward function encodes the robot's general goals.

However, successful application of POMDPs to this domain is probably the most challenging of all the application areas and there currently exists a wide gap between the current state-of-the-art in POMDP research and what is required for a successful autonomous robotic application.

Deriving control rules for robots using POMDP models seems to be best done using models that are at a higher level of abstraction than what the robot actuators and sensors provide (Simmons & Koenig 1995; Cassandra, Kaelbling, & Kurien 1996; Nourbakhsh, Powers, & Birchfield Summer 1995). Another possibility for applying POMDP models closer to the hardware level would be to use a hierarchical arrangement of POMDP models, though this approach is mostly unexplored.

Behavioral Ecology

One can use decision models in the study of the behavior of organisms (Mangel & Clark 1988). As mentioned in the discussion about autonomous robots, no matter the quantity or quality of the sensors, an autonomous agent (i.e., the organism) must still deal with only partial information about its environment.

The premise of this work is that the organism is behaving optimally, given the internal model of the world it is using. The task is to understand exactly what this model is, or to understand the behavior of the organism by identifying the elements the organism chooses to use in its decisions.[1] The researcher will build a model of the states, actions, observations and rewards it believes are important to the organism, compute the optimal policy and then compare the behavior predicted by this policy with the actual behavior of the organism. Using discrepancies, alternate theories are explored, the model is refined, and the process is repeated. This iterative process leads to a better understanding of the organism's behavior and its interaction with the environment.

Machine Vision

Decision process models have been applied to machine vision problems, particularly as it applies to visual attention (Bandera *et al.* 1996). A large body of the work in machine vision uses static images, with uniform resolutions over the entire image and results in algorithms which are computational demanding. To alleviate some of the problems with these approaches, taking their cue from nature, researchers have experimented with the idea of having a small, high resolution area (i.e., a fovea) with a larger, lower resolution area surrounding it. This reduced area now means that it is important to make good decisions about where to focus the attention of the fovea.

The low resolution area, not to mention all the areas for which no part of the visual system is focused, leads naturally to partial observability of the surrounding environment. Because building a POMDP model powerful enough to be used by a general visual system is currently highly impractical. POMDP models are best employed in special purpose visual systems, where the domain has been restricted.

One such special-purpose system exists in the area of gesture recognition (Darrell & Pentland 1996). Here the visual system is capable of tracking the head and hands of a person, where the camera movements between them define the actions in the model. The states of the model correspond to the states of the person or environment; e.g., their hand position, facial expression, etc. Since the visual system cannot focus on all parts of the person, its observations consist of the feedback it gets for the particular area it is currently focused on. The goal is to recognize a particular pattern of gestures or expressions.

[1] We take no position on what it means for an organism to "decide".

Business Applications

Network Troubleshooting

As a network troubleshooting example, consider a large, interconnected electrical distribution network (Thiebeaux *et al.* 1996). When a component fails or a circuit breaker trips, large areas of the population can be affected. Restoring electricity to as many people as quickly as possible is a crucial task that currently takes a high degree of skill and experience. Sending a team out in the field to discover and fix the problem is both time and cost intensive.

The interconnected nature of the network and the presence of remotely controlled switches with remotely accessible circuit-breaker position sensors can allow both reconfiguration and troubling-shooting of the network from the main control station. In fact, the two are intimately tied together; by setting different configurations of the network and monitoring exactly which circuit-breakers are tripped, the controller can localize the faulty component remotely before sending a repair team. However, aside from wanting to locate the failure quickly, there is the competing goal of maintaining service to as many customers as possible.

The partial observability in this domain comes from the limited number of circuit-breaker position sensors and the fact that they do not always provide accurate information. The states are the possible configurations of the network and the possible states of all the components in the network. The actions of the model come from the network of remotely operated electrical switches. The immediate reward would have to be some qualitative estimate of the costs for maintaining customers and the benefits of locating the fault quickly.

Many other types of networks share similar characteristics with the electrical distribution networks; e.g., phone networks, computer networks, etc. They are prone to failures and it is crucial to reconfigure the network or re-route messages while trying to isolate and repair the faulty component. With communication networks, even without faulty components, near or higher than capacity demand can stress a system so that making good routing decisions becomes a crucial goal. Although the details of the model will differ from system to system, complete observability of the state of the entire network is a rare commodity.

Distributed Database Queries

With the explosive expansion of computer networks and the amount of data available on them, more and more researchers are developing distributed information retrieval systems (Bayardo *et al.* 1997). Since an information request is essentially a database query, an important component of such systems is the manner in which queries are processed. Control policies for distributed queries are useful for locating the appropriate sources of the information as quickly or as cheaply as possible. The situation gets more interesting when one considers the information could be duplicated at multiple sources. Naturally, decision process models provide a good formalism for developing query distribution control policies (Segall 1976).

Like the network examples previously presented, it is not feasible to maintain this global system state, requiring either too much hardware or too much network traffic. However, the network traffic as a whole and/or isolated packets do provide some observations about the current network state.

For a POMDP model, the state would be the states of the information sources (up, down, lightly loaded, heavily loaded, etc.) and the states of the network components. The actions would be the various information sources where the query could be submitted. The observations could be results of previous queries, general network traffic, or specific network messages. The immediate rewards could range from costs for accessing information sources to preferences over the response times for the query results.

Marketing

Few products have universal applicability and appeal. Most are targeted toward particular demographics. Sometimes the target audience is more easily identified than others, but more often than not it requires some exploration before a given individual can be identified as either a potential customer or a waste of a salesperson's time. Thus, the person's state is only partially observable, and it requires specific actions to be taken and responses noted before the salesperson decides whether to continue or move on. Application of the POMDP model in this domain, and deriving optimal policies can result in a more effective use of a company's marketing resources.

Building a POMDP model of the internal state of a customer would seem to be equated to modeling a human, but it can be made much more rudimentary than this. At the simplest level, you could have a two state model: `fits-demographics` and `does-not-fit`. More elaborately you can break the notion of fitting the demographics into a number of related attributes, where the more attributes of the demographics the person has the more likely they are to buy the product.

The actions could be questions the salesperson asks the potential customer, or other any other thing the salesperson can do that would requires a response from the customer. The observations are the specific customer reactions to the salesperson or simply how the customer behaves over a period of time. Additionally, the purchasing history of the individual can also be viewed as observations about their state. The immediate rewards for the model will naturally be positive for selling a product and negative for spending time on people that do not fit the particular demographics.

Even with products that have a more universal appeal, personality types will vary as to what marketing techniques are most effective on them. Having a marketing strategy that can take observations from the potential customer to tailor the sales pitch to its most effective form. As companies build up databases of customers and customer profiles, the POMDP models can be refined, honed and made more elaborate over time, thereby increasing the effectiveness of the marketing.

Another interesting marketing application comes from looking at the previous problem from the inside out: when there are a lot of product types and the company wants to focus their marketing resources on the products the potential customer would be most interested in.

As a simple examples, consider a Internet-based company that has many products. The POMDP model used by the company has states consisting of a number of demographic attributes. The products it sells have been categorized according to their profit margin and demographic appeal for each attribute. As customers browse the company's web site, the company can monitor the types of products or pages they seem to be visiting. They can use these observations to tailor the pages (the actions of the model), adding related advertisements and/or links to products and information that matches the demographics of the individual visiting the site.

Questionnaire Design

Somewhat related to the marketing application, is the problem of determining the proper sequencing of questions for a questionnaire where the answers given may be less than truthful (White 1976). Like the marketing example, the states correspond to the type of person being queried. The actions are the specific question set available, the observations are the person response and the objective is to get the proper information as defined by the specific purpose of the questionnaire.

Corporate Policy

Corporate organization and policy are other areas where POMDPs can be applied. From performing internal audits (Hughes 1977) to cost control in accounting (Kaplan 1969), the observed outcomes of the specific corporate actions will never tell the complete story or the current state of the organization. Having a model of the organization allows analysis of the entire structure while also providing optimal business policies for maintaining a stable, productive organization.

Military Applications

The military domain provides many rich and varied applications for POMDPs. The world itself is large and not completely observable. Antagonists expend a lot of effort to ensure their opposition knows as little about them as possible. The military must often deal with both the complications of the world and actions of antagonists at the same time. We highlight a few of the more obvious applications of POMDPs though many others exist.

Moving Target Search

Submarine warfare is one of the clearer examples of the need to search for moving targets, though the problem of locating mobile missile platforms poses similar problems (Eagle 1984; Pollock 1970). Knowledge of the locations of the opposition's assets is always a crucial element in military strategy, but the relative stealthiness of submarines makes they particularly difficult to track.

A POMDP model would have a state space consisting of the possible locations of the opposition's submarine(s). Periodically, reconnaissance aircraft, satellite imagery, sonic buoys, surface vessels, etc. may provide brief observations about the location of a submarine. The state transitions for the submarine's subsequent movements could be based upon terrain, weather, supply lines, etc. Since the goal is to keep as good an estimate of its location as possible, the action of the model are the active measures available to try to detect the submarine. These could be selection of active vs. passive sonar, satellite camera repositioning, reconnaissance flights, helicopter-based sonar sweeps, etc.

Search and Rescue

Similar to the moving target search, is a search and rescue mission. Though there isn't usually the adversarial relationship, there is still the need to develop a search policy, which will be based upon some partial information about the location of the object of the search. In this case, the actions consist of where to deploy the resources so as to maximize the likelihood of finding the object. Observations would be radio transmissions, reports, results of searches in other areas, etc. Time and resources being scarce commodities in a search and rescue mission could greatly benefit from improved control policies.

Target Identification

Detecting the difference between an approaching hostile aircraft and a non-threatening aircraft is not always an easy task. The mis-classification either way could have dire consequences. Although radar and radio transmissions can provide some information about aircraft, they do not always uniquely identify the object, especially when a hostile aircraft is actively trying to disrupt these detection systems. The inaccuracies of the sensors and the uncertainty of the aircraft's movements makes this a very compelling application area for POMDPs.

In building a POMDP model in this domain, the states would be some attributes about the approaching aircraft; aircraft type, altitude, friendly vs. enemy, etc (D'Ambrosio & Fung 1996). The actions here could include active or passive radar, launching intercept aircraft or launching surface-to-air missiles. The observations would be the results of the sensing actions and perceived outcomes of the other actions.

This domain is particularly interesting because the selection between active and passive radar involves a crucial trade-off. On the one hand, active radar will be more accurate and provide more information. However, this also makes the location highly visible to the opposition. Passive radar reduces the chances of having the location exposed to the opposition, but is more error prove in identification. By incorporating the relative level of visibility into the state space, and defining how each sensor increases the visibility, the optimal POMDP policy would make the best trade-offs between these competing goals.

Weapon Allocation

An attack aircraft has a maximum payload it can carry and policies for how best to use its assets becomes a crucial element in a confrontation. Although a lot of planning goes into the exact targets and armaments needed before an aircraft ever begins its mission, there would be a great benefit for being able to adjust the mission in reaction to unfolding events that occur after the aircraft becomes airborne.

Aside from actually releasing the weapons, aircraft usually will record the results so that they can later be analyzed to determine the effectiveness of the attack and whether or not follow-up missions will be necessary; i.e. bomb damage assessment (BDA).

Consider the situation of an aircraft departing with two air-to-surface missiles and two targets. After releasing a weapon at the first target and recording the results, the aircraft would normally proceed to the next target. However, suppose the first target had a much higher priority than the second, then if the first weapon did not produce the desired damage, a better choice might be for the aircraft to use its second missile on the first target again.

The recordings taken of the damage are by nature never totally conclusive, which makes this very much of a partially observable problem (Yost 1998). By modeling the level of damage as states of a POMDP and defining some rough relationship between the state and the observations the recording provide, we could derive better, more reactive policies that account for the relative priorities of targets and costs for attacking them.

Social Applications

Education

Somewhat related to the marketing application, is the development of teaching strategies. The relationship comes from trying to model the internal mental state of an individual, only this time, with the more noble goal of trying to find the best way to teach concepts, rather than sell products.

Consider the task of trying to teach n concepts to an individual (Karush & Dear 1967; Smallwood 1971). In a POMDP formulation, the state of the student could have a boolean attribute for each of the concepts indicating whether it had been learned or not. The actions available to the teacher would be various types of learning techniques for each of the concepts, potentially with some teaching actions representing an integration of concepts. The observations in this model would be the results of tests given periodically. One could even include the testing as an action choice in the model, so that testing was done when it would be most beneficial for the teacher. The goal could be to teach as many of the concepts in a finite amount of time, or to minimize the time required to learn all the concepts. If time is finite, the concepts could be prioritized, which would be reflected in the immediate reward structure of the model.

This could even be made more elaborate if you incorporate state's of the student corresponding to the way they learn best: e.g., tactile, auditory, verbal, etc. The actions available to the teacher could also be broken down along this dimension and, as the teaching process continued, the optimal policy based on the test observations should steer the action choices to those more closely matched to their learning style.

Medical Diagnosis

Medical diagnosis, while quite advanced, is still a task that requires a significant amount of skill and experience, and which is still prone to errors. The difficultly for the physician is that the exact internal state of the patient never completely reveals itself (Hauskrecht 1997). In an attempt to diagnose a patient, the physician has a number of actions available: conduct laboratory tests, prescribe drug treatments, perform exploratory surgery or recommend various forms of physical therapy. Each of these incurs a cost to the patient, both in monetary units and in terms of health risks. Additionally, each will result in the physician getting observations as results of the various actions chosen for the patient. Here again we see the trade-offs becoming crucial, in this case between the health-risks/costs of procedures and the accuracy of the information they provide.

Health Care Policymaking

If one pushes the medical diagnosis problem up from the level of individuals and physicians to groups of individuals and policymakers, we see that here too, models that incorporate partial observability are useful (Smallwood, Sondik, & Offensend 1971). The states are now the health states pertaining to individual populations and the actions are policy decisions such as how much and where funds should be spent: research, immunization programs, educational programs, awareness advertising, etc. Again, observations can be made on the populations about their health, but this is only probabilistically related to the true underlying state of health.

A crucial trade-off that needs to be made in these systems is the short-term versus long-term gains that the different action choices yield. Building models of such systems and finding optimal policies for them can greatly help the policy-makers reason about the systems they are trying to manage.

Limitations

Despite the relatively broad range of application areas shown in this paper, POMDPs do have limitations and do not easily handle problems with certain characteristics. In particular, our model assumed finite sets for the states, actions, observations and required the decision points to occur at discrete time steps. Although any continuous space or time can be discretized, this is not always the best approach for some applications. Additionally, implicit in all the MDP models is the assumption that the process being modeled actually obeys the Markov assumption.

The other problem with the POMDP model is that it is data intensive. It requires that every transition probability, every observation probability and every immediate reward be specified for each state, action and observation. The first question is whether or not the application domain can provide all of this information, and the second is more of a user interface issue about how to provide a means to build such models. Another research area tries to address the former issue by using data gathered from observations or simulations of the process to learn or adjust the parameters of the model (Koenig & Simmons 1996; Shatkay & Kaelbling 1997). There are also model-free techniques for finding control policies for Markov models (Bertsekas & Tsitsiklis 1996).

The discussion above is mostly concerning the theoretical limitations of the model. For most of the application areas described, the real limitations currently lie in the representational and computational areas. In many of the applications discussed, we refer to representing states as a series of attributes. Most POMDP algorithms would require every attribute value combination to be enumerated, which can cause problems that are conceptually small to require large state spaces. Since problems are more naturally and easily specified in this factored, attribute form, a rich research area is to develop effective algorithms that can use this representation (Draper, Hanks, & Weld 1993; Boutilier & Poole 1996).

Computationally, finding the optimal policy for a general POMDP is intractable (Papadimitriou & Tsitsiklis 1987). Approaches to dealing with this are to to be satisfied with less than optimal solutions or to develop algorithms that can exploit problem characteristics (Littman, Cassandra, & Kaelbling 1995; Parr & Russell 1995; Zhang & Liu 1996; 1997; Castanon 1997; Hauskrecht 1997; Brafman 1997; Cassandra 1998). For the latter, applications can help drive the search for useful characteristics that can be exploited.

References

Bandera, C.; Vico, F. J.; Bravo, J. M.; Harmon, M. E.; and III, L. C. B. 1996. Residual Q-Learning applied to visual attention. *Machine Learning* 20–27.

Bayardo, R.; Bohrer, W.; Brice, R.; Cichocki, A.; Fowler, J.; Helal, A.; Kashyap, V.; Ksiezyk, T.; Martin, G.; Nodine, M.; Rashid, M.; Rusinkiewicz, M.; Shea, R.; Unnikrishnan, C.; Unruh, A.; and Woelk, D. 1997. Infosleuth: An agent-based semantic integration of information in open and dynamic environments. *Proceedings of the ACM SIGMOD International Conference on Management of Data.*

Bertsekas, D. P., and Tsitsiklis, J. N. 1996. *Neuro-Dynamic Programming.* Belmont, Massachusetts: Athena Scientific.

Boutilier, C., and Poole, D. 1996. Computing optimal policies for partially observable decision processes using compact representations. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, 1168–1175.

Brafman, R. I. 1997. A heuristic variable grid solution method for POMDPs. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence*, 727–733.

Cassandra, A.; Kaelbling, L.; and Kurien, J. 1996. Acting under uncertainty: Discrete bayesian models for mobile-robot navigation. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems.*

Cassandra, A. R.; Kaelbling, L. P.; and Littman, M. L. 1994. Acting optimally in partially observable stochastic domains. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, volume 2, 1023–1028. Menlo Park, California: American Association for Artificial Intelligence.

Cassandra, A. R. 1998. *Exact and Approximate Algorithms for Partially Observable Markov Decision Processes.* Ph.D. Dissertation, Department of Computer Science, Brown University.

Castanon, D. A. 1997. Approximate dynamic programming for sensor management. In *Proceedings of the Conference on Decision and Control* To appear.

Crites, R. H. 1996. *Large-scale Dynamic Optimization Using Teams of Reinforcement Learning Agents.* Ph.D. Dissertation, University of Massachusetts.

D'Ambrosio, B., and Fung, R. 1996. Far sighted approaches to sensor management experiments in reinforcement learning. Technical report, Prevision.

Darrell, T., and Pentland, A. 1996. Active gesture recognition using partially observable Markov decision processes. In *Thirteenth IEEE International Conference on Pattern Recognition (IPCR'96).* Also appears as Technical Report No. 367 from M.I.T. Media Laboratory Perceptual Computing Section.

Draper, D.; Hanks, S.; and Weld, D. 1993. Probabilistic planning with information gathering and contingent execution. Technical Report 93-12-04, University of Washington.

Eagle, J. N. 1984. The optimal search for a moving target when the search path is constrained. *Operations Research* 32(5):1107–1115.

Eckles, J. E. 1968. Optimum maintenance with incomplete information. *Operations Research* 16:1058–1067.

Ellis, H.; Jiang, M.; and Corotis, R. B. 1995. Inspection, maintenance, and repair with partial observability. *Journal of Infrastructure Systems* 1(2):92–99.

Hauskrecht, M. 1997. *Planning and Control in Stochastic Domains with Imperfect Information.* Ph.D. Dissertation, Massachusetts Institute of Technology.

Hughes, J. 1977. Optimal internal audit timing. *Accounting Review* LII:56–58.

Kaplan, R. 1969. Optimal investigation strategies with imperfect information. *Journal of Accounting Research* 7:32–43.

Karush, W., and Dear, R. 1967. Optimal strategy for item presentation in learning models. *Management Science* 13:773–785.

Koenig, S., and Simmons, R. 1996. Unsupervised learning of probabilistic models for robot navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation*.

Lane, D. E. 1989. A partially observable model of decision making by fishermen. *Operations Research* 37:240.

Littman, M.; Cassandra, A.; and Kaelbling, L. 1995. Learning policies for partially observable environments: Scaling up. In Prieditis, A., and Russell, S., eds., *Proceedings of the Twelfth International Conference on Machine Learning*, 362–370. San Francisco, CA: Morgan Kaufmann. Reprinted in *Readings in Agents*, Michael H. Huhns and Munindar Singh, eds., 1997.

Lovejoy, W. S. 1991. A survey of algorithmic methods for partially observed Markov decision processes. *Annals of Operations Research* 28(1):47–65.

Mangel, M., and Clark, C. W. 1988. *Dynamic Modeling in Behavioral Ecology.* Princeton, New Jersey: Princeton University Press.

Nourbakhsh, I.; Powers, R.; and Birchfield, S. Summer 1995. Dervish: An office-navigating robot. *AI Magazine* 53–60.

Papadimitriou, C. H., and Tsitsiklis, J. N. 1987. The complexity of Markov decision processes. *Mathematics of Operations Research* 12(3):441–450.

Parr, R., and Russell, S. 1995. Approximating optimal policies for partially observable stochastic domains. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 1088–1094. Morgan Kaufmann.

Pierskalla, W., and Voelker, J. 1976. A survey of maintenance models: The control and surveillance of deteriorating systems. *Naval Research Logistics Quarterly* 23:353–388.

Pollock. 1970. A simple model of search for a moving target. *Operations Research* 18:883–903.

Puterman, M. L. 1994. *Markov Decision Processes — Discrete Stochastic Dynamic Programming.* New York, New York: John Wiley & Sons, Inc.

Rosenfeld, D. 1976. Markovian deterioration with uncertain information. *Operations Research* 24(1):141–155.

Ross, S. M. 1971. Quality control under Markovian deterioration. *Management Science* 17(9):587596.

Segall, A. 1976. Dynamic file assignment in a computer network. *IEEE Transactions on Automatic Control* AC-21:161–173.

Shatkay, H., and Kaelbling, L. P. 1997. Learning topological maps with weak local odometric information. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence*.

Simmons, R., and Koenig, S. 1995. Probabilistic navigation in partially observable environments. In *Fourteenth International Joint Conference on Artificial Intelligence*, 1080–1087. Montreal, Canada: Morgan Kaufmann.

Smallwood, R. D., and Sondik, E. J. 1973. The optimal control of partially observable Markov processes over a finite horizon. *Operations Research* 21:1071–1088.

Smallwood, R.; Sondik, E.; and Offensend, F. 1971. Toward and integrated methodology for the analysis of health-care systems. *Operations Research* 19:1300–1322.

Smallwood, R. 1971. The analysis of economic teaching strategies for a simple learning model. *Journal of Math Psych* 8:285–301.

Thiebeaux, S.; Cordier, M.-O.; Jehl, O.; and Krivine, J.-P. 1996. Supply restoration in power distribution systems — a case study in integrating model-based diagnosis and repair planning. In *Proceedings of the Twelfth Annual Conference on Uncertainty in Artificial Intelligence (UAI–96)*, 525–532.

White, III, C. C. 1976. Optimal diagnostic questionaires which allow less than truthful responses. *Information and Control* 32:61–74.

White, III, C. C. 1977. A Markov quality control process subject to partial observation. *Management Science* 23(8):843–852.

White, III, C. C. 1979. Bounds on optimal cost for a replacement problem with partial observations. *Naval Research Logistics Quarterly* 26:415–422.

White, III, C. C. 1991. Partially observed Markov decision processes: A survey. *Annals of Operations Research* 32.

Yost, K. A. 1998. Solution of sequential weapons allocation problems with imperfect, costly information. Presented at the Sixth INFORMS Computer Science Technical Section Conference.

Zhang, N. L., and Liu, W. 1996. Planning in stochastic domains: Problem characteristics and approximation. Technical Report HKUST-CS96-31, Department of Computer Science, Hong Kong University of Science and Technology.

Zhang, N. L., and Liu, W. 1997. Region-based approximations for planning in stochastic domains. In

Proceedings of the Thirteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI–97), 472–480.