

Boğaziçi University
Department of Computer Engineering
Proposal for Ms. Thesis

**Planing and Acting In Stochastic Domains
For Mobile Robot Control**

M. Toygar KARADENİZ
Advisor : Assoc. Prof. H. Levent Akın

1997

Motivation

The proposed study will be mainly about mobile robot control planning and acting in stochastic domains. For this purpose, we are going to use POMDP (partially observable Markov decision processes) to formalize the position of the robot respect to the environment. The foundations of POMDPs are derived from Markov decision problems (MDPs in short). MDPs are generally useful models for planning and acting in stochastic domains. Although general structure of MDPs is very well studied, there are two extensions which may be the subject of interest. First is enlarging the state space which is the collection of possible robot positioning to include many distinct probabilities. This will absolutely increase the complexity. Second is to keep up with the partially observable environments.

In MDP models, the agent (robot) is assumed to be able to observe exactly what state the environment is in at all times. So, these models are always deterministic. In POMDP models, this assumption is not valid. The agent (robot) when taking steps in the environment, makes observations of the environment and tries to determine the environmental state. But this is usually a tough work, because these observations are generally noisy. The way around for the agent (robot) is to convert these observations into belief states which encode the agent's information about the current state of the environment. Then, it must behave with this knowledge.

The study will be concerned with the following steps :

- investigating the current state-of-art for POMDP studies used for controlling robots (a survey)
- examining the algorithms proposed to attack POMDP problems of mobile robotics
- preparing a test-bed for comparison of the algorithms when used in the mobile robot domain
- determine the weak points of the examined algorithms
- do a study to, hopefully, propose a new algorithm to cover these weaknesses

Scope

The scope of this study is the generic environment with one mobile robot in it and the relevant conversion of it to a POMDP. The goal, here, is to find an optimal policy. A policy is a strategy for selecting robot actions based on the information known by the agent (robot). Moreover, that strategy has to maximize an infinite-horizon, discounted optimality criterion.

A POMDP is a tuple $\langle S, A, T, R, O, \Omega \rangle$ where S is a set of states, A is a set of actions and Ω is a set of observations. Our interest will only be with the case in which these sets are finite. The functions T and R define a Markov decision process

(MDP) with which the agent interacts without direct information as to the current state. The T transition function specifies how the various actions affect the state of the environment, $T : S \times A \Rightarrow \Pi(S)$, where Π represents the set of discrete probability distributions over a finite set. R is the collection of immediate rewards and is defined as $R : S \times A \Rightarrow R$. The agent decisions are made based on information coming from its sensors. So, $O : S \times A \Rightarrow \Pi(\Omega)$.

Methodology

State-of-Art for POMDP Studies

In the very first part of this study, we will be doing a complete survey about the current state-of-art for POMDP studies done on mobile robot domains. This will, surely, be a knowledge collection part of the study and the collected knowledge will be used in the further steps. The checkpoint about this phase is going to be the preparation of a survey report. This report will also be added to the final study report as a first chapter. The survey sources can be determined beforehand as the libraries, books, papers, the Internet etc.

Examining Proposed Algorithms

Converting an robot environment to its relevant POMDP representation, is the easy part. The more challenging one is finding an algorithm to solve that POMDP and show that the environment is fully learned by the robot. A variety of algorithms have been developed for solving POMDP problems, but as the problems are known to be computationally challenging, most techniques turned to be very inefficient to be used on all but the smallest problems. So, the generality and expressiveness power of POMDPs has a cost. Only very small problems can be solved by using the available techniques. The following is the list of the current POMDP algorithms which are going to be examined in this study :

- Truncated Exact Value Iteration Algorithm (Witness Algorithm)
- The Q_{MDP} Value Method
- Replicated Q-Learning
- Linear Q-Learning

Preparing a Test-Bed

Next phase of this proposed study will be the development of a test-bed system for the examined algorithms to determine the power of each one in robot domains. The mentioned test-bed will, surely, be a software program working under possibly Windows-like operating system. This choice is mainly because of the visual purposes. The test-bed software is planned to have the following components :

- **Visual Component** : During the development and testing phases, visuality is needed to get the practical idea about the power of the algorithm. This component will surely be a representation of the environment and the position of the agent (robot) respect to the environment. By the help of it, the steps that are taken by the agent (robot), will be examined visually and on-line.
- **Trace Component** : As we are going to make some kind of comparison between the algorithms in this study, we may need to take some snapshots and compare the results at these checkpoints. So, a component to do this job is necessary. The snapshots generated by this component may contain many valuable results like the amount of memory source used, the time complexity, the summary of steps taken etc.
- **Algorithm Component** : All POMDP algorithms will be saved as a packet in the algorithm component of this system. Actually, algorithm component is a passive entity, but also the center of the operations.
- **Example Component** : Every test system needs realistic examples to test the concepts extensively and closely to reality. Surely, we need such a component in our system, too. The main job of this component will be generating environmental states and their POMDP tuples as $\langle S, A, T, R, O, \Omega \rangle$ in order to be attacked by the algorithm set. Moreover, the size of the generated POMDPs must be given by an outside input. So, the algorithm power versus problem size can easily be determined.
- **Timing Component** : This is the supervisor component that does the management jobs concerning all the other components of the system. Actually, the timing component is the main module.

Of course, this system is just a proposal in the mean time. That means when necessary additions and/or modifications to the proposed components come into play, they may be done.

Possibly A New Algorithm

As mentioned, there are a variety of algorithms developed for attacking POMDP problems especially when used in intelligent robot environment learning

systems. But, they all have some shortcomings or deficiencies. Many of them suffer from the exponential explosion of the number of possible states and somewhat like problems. So, shortly, there is no one best algorithm for every case.

Final phase of this study is being planned to be a proposal of a new algorithm which is trying to cover all the shortcomings of the mentioned algorithms. Of course, the outline of this algorithm is not clear at this premature point. But, at the end of this study, any new algorithm, if there is one which is enough important to mention, will be presented. Or if there is none, this fact will be proven. Moreover, this new algorithm will be included in the comparison and the results concerning these will be given by using numerical tables and charts. Certainly, the pseudo-code will be included, too.

Tentative Schedule

- **February - March 1997** :_Determining state-of-art for POMDPs.
- **May - June 1997** : Examining proposed algorithms.
- **October - December 1997** : Preparing a test-bed.
- **December - January 1998** : Constructing possibly a new algorithm.

References

- M. L. Littman, A. R. Cassandra and L. P. Kaelbling, "Learning policies for partially observable environments : Scaling up," *Proceedings of the Twelfth International Conference on Machine Learning*, 1995.
- A. R. Cassandra, J. A. Kurien and L. P. Kaelbling, "Acting Under Uncertainty : Discrete Bayesian Models for Mobile-Robot Navigation," *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1996.
- M. L. Littman, T. L. Dean and L. P. Kaelbling, "On the Complexity of Solving Markov Decision Problems," *Proceedings of the Eleventh International Conference on Uncertainty in Artificial Intelligence*, 1995.
- T. Dean, L. P. Kaelbling, J. Kirman and A. Nicholson, "Planning Under Time Constraints in Stochastic Domains," *Artificial Intelligence*, Vol. 76, 1995.
- A. P. Duchon, W. H. Warren and L. P. Kaelbling, "Controlling Behavior with Optical Flow," *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*, 1995.

