Paper Review

Title: ACTING OPTIMALLY IN PARTIALLY OBSERVABLE STOCHASTIC DOMAINS

by

Anurag Saraswat (M20CS066)

Date: April 5, 2021

1 Abstract

Review of paper titled ”Acting Optimally in Partially Observable Stochastic Domains” is presented. Review is initiated by defining a problem presented in paper and then discussing the importance of the problem with an example of robot planning problem scenario. In the discussion part, the algorithm presented by the author is discussed by breaking the algorithm into its constituent parts. Further there is discussion on various missing information that are not provided by the author and cannot be inferred from provided information. At the end of discussion critical comments on claims made by the author are discussed. Review is concluded by summarizing key ideas and discussion on the possibility of further extension of concepts presented in paper and applications lies under problem domain.

2 Introduction

Partial observable markov decision process (POMDPs) model environment in which agent does not have complete information about the state of the environment i.e environment is only partially observable.[2] In such a type of environment an agent works over beliefs state space to perform action. In this paper the author extends concepts of Markov Decision Process models to partial observable markov decision processes (POMDP). An example of working in a POMDPs environment is planning problems in robotics. Consider an instance when a robot is performing some task and suddenly stops. Now , the robot has to find out from where it continues his task. At this point the robot calibrates itself by going to some known position using its belief and then continues its task.

3 Discussion

In this paper the author extends concepts of Markov Decision Process models to partial observable markov decision processes (POMDP) and enumerates ”Witness Algorithm” which will be able to find out optimal policy for agents in a partially observable environment. Witness algorithm starts by initializing set of vectors which represents Vt* given Vt-1 and is updated iteratively like in the bellman equation. Here , V_{t-1} given belief b is represented by max of α in V_t over product of b.α . Here, b is belief and α is known to be witness. And using witness we decide when to stop iteration.

Proposed witness algorithm can be broken down into two steps. In the first step we partition the continuous space iteratively into small regions which will share the same utility or belief/observability or have related action. In step two we construct a policy graph using these small related partitioned regions. Policy graph is therefore a kind of representation of optimal policy. Initially agent chooses action associated with starting node and then from starting depending on observation agent makes transition to other states. In policy graph current state is capable to give information about past experience of agent and also capable enough to make future decisions. New information / observation is incorporated in decision making through arcs of graph. Policy graphs are not only efficient but also simple to execute. In nutshell we can conclude from two steps discussed above that the algorithm first presented the problem of POMDP as continuous space MDP problem and then solved the problem in domain of MDP by applying value iteration to select action which will give the maximum expected value function using beliefs.

Value function we discussed above is found to be convex and it is approximated by sets of vectors. Value is calculated by a scalar product between these sets of vectors and belief states gives maximal value and is used for value function. The process of transforming these vectors associated to value function to policy graph is not mentioned in paper.

Paper lacks some statistics which could make the author's claims more fair like convergence analysis of algorithms. Also author claim the approach to be better than the previous approach but only observational intuition is provided no result is included in paper. Detailed comparison with other algorithms must be provided to support the claim.

4 Conclusions

The content of the paper is precise and gives an overview of POMDP and how to transform POMDP to MDP using belief or observation. Also, the author presented an algorithm to solve the POMDP decision problem efficiently. The results presented in paper are limited and there is immense scope to extend work further. Like instead of searching in value space we can extend algorithm to search in policy space which could even turn out to be more efficient than value iteration. Also there is one more possible direction to optimize further by limiting search space.[1] This can be achieved by pruning or using some heuristics. Defined approaches can be utilized to solve various problems like robot planning , surveillance and tracking.

References

[1] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra. Partially observable markov decision processes for artificial intelligence. In L. Dorst, M. van Lambalgen, and F. Voorbraak, editors, *Reasoning with Uncertainty in Robotics*, pages 146–163, Berlin, Heidelberg, 1996. Springer Berlin Heidelberg.

[2] K. Åström. Optimal control of markov processes with incomplete state information. *Journal of Mathematical Analysis and Applications*, 10(1):174–205, 1965.