Spatial and Temporal Abstractions in POMDPS: Learning and Planning

Georgios Theocharous and leslie Pack Kaelbling

Introduction: A popular approach to artificial intelligence is to model an agent and its interaction with its environment through actions, perceptions, and rewards [1]. Intelligent agents should choose actions after every perception, such that their long-term reward is maximized. A well defined framework for this interaction is the partially observable Markov decision process (POMDP) model. Unfortunately, standard methods for planning, inference, and learning with POMDPs take time at least exponential in the number of (discrete) states, S, making them impractical for large problems.

Recent research has explored the advantages of spatial and temporal abstraction in POMDPs [2]. A hierarchical extension to POMDPs such as H-POMDPs, [5] represents the state-space at multiple levels of abstraction, and scales much better to larger environments. In particular, it simplifies the planning and learning problems. Planning is simpler (requires less time) in H-POMDPs because abstract states (at the coarse-level of resolution) have lower entropy, i.e., are more deterministic [2]. Learning is simpler (requires less data) in H-POMDPs because the number of free parameters is reduced, and the structure of the model provides a way of encoding prior knowledge.

We are continuing our exploration of spatio-temporal abstractions for POMDPs. In particular in learning we investigate how to represent H-POMDPs as dynamical Bayesian networks [4]. In planning, we describe a new reinforcement learning algorithm over belief states, which uses macro-actions [3].

Learning: In learning we explore the advantages of representing H-POMDPs as dynamic Bayesian networks (DBNs). In particular, we focus on the special case of using H-POMDPs to represent multi-resolution spatial maps for indoor robot navigation. Our results show that a DBN representation of H-POMDPs can train significantly faster than the original learning algorithm for H-POMDPs or the equivalent flat POMDP, and requires much less data. In addition, the DBN formulation can easily be extended to parameter tying and factoring of variables, which further reduces the time and sample complexity. This enables us to apply H-POMDP methods to much larger problems than previously possible. Figure 1 depicts an H-POMDP model and its DBN representation.

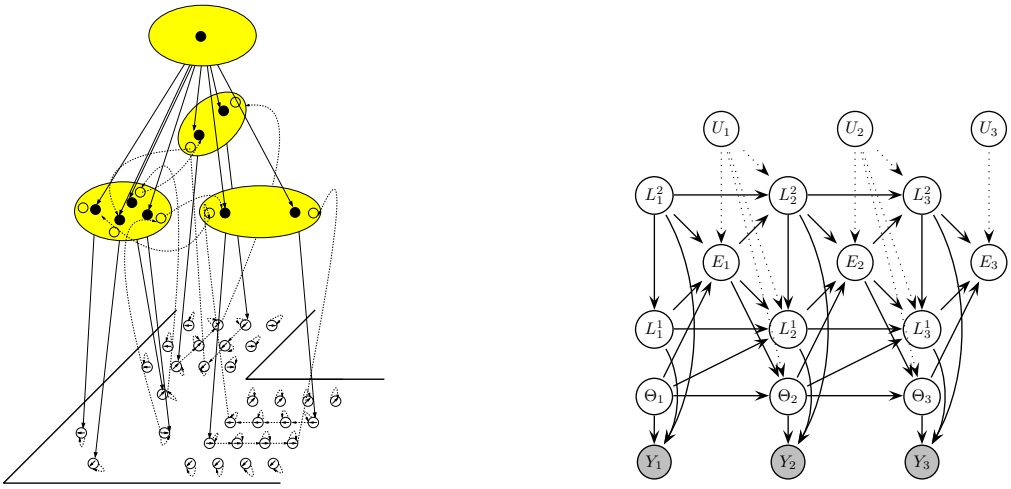

Figure 1: The figure on the left shows a hierarchical POMDP representation of spatial indoor environments. Abstract states (shaded circles) represent corridors, junctions and buildings. Production states (empty circles) represent robot location and orientation. At higher levels of abstraction there is less uncertainty, which results in better robot state estimation. The figure on the right shows a 2-level factored H-POMDP represented as a DBN. The U_t nodes denote the action nodes. The L_t^2 nodes denote the abstract state, the L_t^1 nodes denote the concrete location, the Θ_t nodes denote the orientation, the E_t nodes denote the state of the exit variable, and Y_t denotes the state of the observation variables.

Planning: In planning we explore the fact that useful POMDP solutions do not require consideration of the entire belief space. We extend this idea with the notion of temporal abstraction. We present and explore a new reinforcement learning algorithm over grid-points in belief space, which uses macro-actions

and Monte Carlo updates of the Q-values. We apply the algorithm to a large scale robot navigation task and demonstrate that with temporal abstraction we can consider an even smaller part of the belief space, we can learn POMDP policies faster, and we can do information gathering more efficiently.

In a regular grid-based approach, we discretize the belief space by covering it with a uniformly-spaced grid as shown in Figure 2, then solve an MDP that takes those grid points as states. Unfortunately, the number of grid points required rises exponentially in the number of dimensions in the belief space, which corresponds to the number of states in the original space. In our work, we allocate grid points from a uniformly-spaced grid dynamically by simulating trajectories of the agent through the belief space. At each belief state experienced, we find the grid point that is closest to that belief state and add it to the set of grid points that we explicitly consider. In this way, we develop a set of grid points that is typically a very small subset of the entire possible grid, which is adapted to the parts of the belief space typically inhabited by the agent (see right part of figure 2).

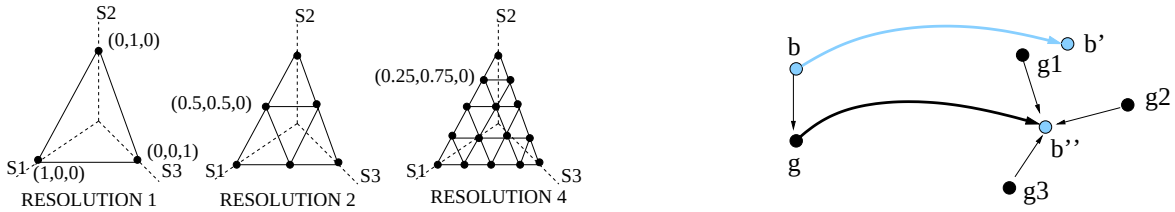

Figure 2: The figure on the left depicts various regular dicretizations of a 3 dimensional belief simplex. The belief-space is the surface of the triangle, while grid points are the intersection of the lines drawn within the triangles. In the right figure an agent finds itself at a belief state b. It maps b to the grid point g, It chooses a macro action and executes it starting from the chosen grid-point, using the primitive actions and observations that it does along the way to update its belief state. It gets a value estimate for the resulting belief state b'' through interpolation from nearby grid points $g1$, $g2$, and $g3$. The agent executes the macro-action from the same grid point g multiple times so that it can approximate the probability distribution over the resulting belief-states b''. Finally, it can update the estimated value of the grid point g and execute the macro-action chosen from the true belief state b. The process repeats from the next true belief state b'.

Conclusion and future directions: In general, spatio-temporal abstractions and multi-resolution representations in POMDPS are necessary in-order to scale up to large domains. We are currently exploring methods for automatically deriving these abstractions for arbitrary POMDP models.

Acknowledgement: This work was supported in part by NASA award # NCC2-1237 and in part by DARPA contract # DABT63-99-1-0012.

References:

[1] S. J. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2nd edition, 2003.

[2] Georgios Theocharous. *Hierarchical Learning and Planning in Partially Observable Markov decision Processes*. PhD thesis, Michigan State University, 2002.

[3] Georgios Theocharous and Leslie Pack Kaelbling. Approximate planning with macro-actions in POMDPs. In *Neural Information Processing systems 16 (NIPS)*, Vancouver, 2003.

[4] Georgios Theocharous, Kevin Murphy, and Leslie Pack Kaelbling. Representing hierarchical POMDPs as DBNs for multi-scale robot localization. In *IEEE conference on Robotics and Automation (ICRA)*, New Orleans, 2004.

[5] Georgios Theocharous, Khashayar Rohanimanesh, and Sridhar Mahadevan. Learning hierarchical partially observable Markov decision processes for robot navigation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Seoul, Korea, 2001.