Grounding State Representations in Sensory Experience

Daniel Nikovski and Illah Nourbakhsh

Problem: Although the world surrounding robots can be described most naturally by means of continuous variables such as position, orientation, temperature, light intensity, etc., autonomous reasoning, planning, and learning in continuous domains has proven to be very difficult. One way humans deal with the overwhelming complexity of the environment is by developing elaborate concise discrete descriptions of the processes they are involved in and using these descriptions to achieve their everyday goals. For example, when moving from one location to another, humans tend to plan their actions in terms of transitions between rooms, corridors, cities, valleys, etc., which are solely the products of their minds and impose a convenient discrete representation over the continuous environment. Furthermore, such descriptions contain state information in cases of perceptual aliasing, when the immediate percepts cannot disambiguate the state of the world. We address the problem of endowing robots with the same ability to create discrete representations of the continuous world around them, with the objective of improving their reasoning and planning performance.

Impact: Our short-term objectives are to develop algorithms for autonomous acquisition of discrete world models by robots and using them for navigation in unknown environments. If successful, these algorithms would allow robots to be deployed in completely unknown environments, which they could explore without human intervention. Our long-term goal is to make it possible for robots to develop their hierarchical internal symbolic representations of the surrounding world and reuse these representations for a wide variety of tasks.

State of the Art: In most robotic applications, a human designer defines manually a world model for the particular domain, in which a robot will be operating. For mobile robots, this model can be a map (either metric or topological); for other domains, a set of STRIPS-style operators that manipulate the state of the world can be defined and used by symbolic planners. The manual creation of these models is one of the largest costs associated with deploying robots.

Currently, there are no general methods for autonomous acquisition of discrete models of robot actions in continuous environments with hidden state. For some domains, such as navigation in office spaces, there exist methods for acquiring maps of the environment, but they are domain-specific and are not likely to be useful in other domains. On the other hand, the field of reinforcement learning has made substantial progress in learning general probabilistic models with hidden state, but research has focused exclusively on discrete worlds such as mazes. Extending the applications to continuous domains has proven very hard, even when the world is fully observable.

Approach: Our approach is to develop algorithms for learning partially-observable Markov decision process (POMDP) models with discrete hidden state and continuous observations. The effect of such algorithms is defining (grounding) discrete hidden states into continuous observations, coming directly from the sensors of a robot. For the purposes of mobile robot navigation, such hidden states correspond roughly to places and areas in the operating space of a robot; in the general case they correspond to the conceptual representations of the outer world, which humans use in reasoning and planning their actions efficiently.

Since POMDP models are simply hidden Markov models (HMM) augmented with actions, it should be possible, at least in theory, to use general methods for learning probabilistic models with hidden state, such as the Baum-Welch algorithm. In practice, however, such algorithms often get stuck in shallow local maxima of data likelihood even when the observations are discrete, which leads to catastrophic performance when the models are used to plan actions [1]. Learning such models with continuous observations is even harder – these algorithms almost never converge to usable models unless their are initialized with emission models, which are pretty close to the final solutions.

In order to overcome these shortcomings of iterative optimization algorithms such as Baum-Welch, we are considering a completely different approach to learning probabilistic models with hidden state, based on merging states. Proposed originally by Stolcke and Omohundro for the purposes of learning HMMs for speech recognition, this approach starts with a completely deterministic, albeit overfit model, and gradually merges states so as to simplify the model and improve its generalization abilities. Still, this algorithm proceeds greedily and never reconsiders suboptimal state mergers, which sometimes results in models of poor quality [2]. To solve this problem, we developed a novel algorithm (SMTC), which considers all possible mergers before carrying out any of them. The algorithm performs clustering in the embedding space of trajectories leading into and out of a particular state, and employes an efficient algoithm for computing the similarities between pairs of trajectories. So far, experimental results in small test worlds have shown superiority of our algorithm over Baum-Welch and the greedy state-merrging algorithm of Stolcke and Omohundro [3]. Table 1 shows results for three learning methods and five planners: AP – assumptive planning; MLS – most-likely state MDP-based; Voting – voting MDP-based; Q_{MDP} – MDP-based proportional to Q-values. Shown are the average reward over five trials, the associated standard deviation, and the statistical z test for difference between the achieved reward and that of random action selection. The last row shows the performance of the planners when the true POMDP model is available to them.

METHOD	AP	MLS	VOTING	Q_{MDP}	RANDOM
BW	$0.32 \pm 0.06/+4.13$	$0.25 \pm 0.11/-1.65$	$0.26 \pm 0.11/-0.63$	$0.26 \pm 0.11/-0.63$	0.269 ± 0.056
BFMM	$0.26 \pm 0.16/-0.40$	$0.23 \pm 0.14/-2.52$	$0.23 \pm 0.14/-2.52$	$0.26 \pm 0.16/-0.40$	0.269 ± 0.056
SMTC	$0.33 \pm 0.07/+4.94$	$0.33 \pm 0.07/+4.94$	$0.33 \pm 0.07/+4.94$	$0.33 \pm 0.07/+4.94$	0.269 ± 0.056
TRUE	$0.65 \pm 0.00/+32.88$	$0.65 \pm 0.00/+32.88$	$0.65 \pm 0.00/+32.88$	$0.65 \pm 0.00/+32.88$	0.269 ± 0.056

Future Work: The current implementation of the algorithm makes impractical learning with a trace longer than 300 observations, and in order to scale up our algorithm to larger worlds, we would have to find more efficient methods for clustering with similarity matrices only. Another direction we are planning to explore are the algorithms for learning probabilistic models with hidden state, suggested by Brand. These methods avoid shallow local maxima in likelihood by means of entropic priors, parameter extinction, and deterministic annealing, and since they have already proven their superioriry over Baum-Welch in areas such as speech and character recognition, it can be expected that they will have equal success in learning POMDP models.

Selected References:

[1] D. Nikovski. Learning stationary temporal probabilistic networks. In *Proceedings of the Conference on Automated Learning and Discovery CONALD'99*. Carnegie Mellon University, Pittsburgh, PA, 1998.

[2] D. Nikovski and I. Nourbakhsh. Learning discrete Bayesian models for autonomous agent navigation. In *Proceedings of the 1999 IEEE International Symposium on Computational Intelligence in Robotics and Automation CIRA'99.*, pages 137–143. IEEE, Monterey, CA, 1999.

[3] D. Nikovski and I. Nourbakhsh. Learning probabilistic models for decision-theoretic navigation of mobile robots. In *Proceedings of the Seventeenth International Conference in Machine Learning ICML'2000*. Stanford, 2000.