

Grounding State Representations in Sensory Experience for Reasoning and Planning by Mobile Robots

Daniel Nikovski

The Robotics Institute, Carnegie Mellon University
5000 Forbes Ave., Pittsburgh, PA 15213
nikovski@cs.cmu.edu

Abstract

We are addressing the problem of learning probabilistic models of the interaction between a mobile robot and its environment and using these models for task planning. This requires modifying the state-of-the-art reinforcement learning algorithms to deal with hidden state and high-dimensional observation spaces of continuous variables. Our approach is to identify hidden states by means of the trajectories leading into and out of them, and perform clustering in this embedding trajectory space in order to compile a partially observable Markov decision process (POMDP) model, which can be used for approximate decision-theoretic planning. The ultimate objective of our work is to develop algorithms that learn POMDP models with discrete hidden states defined (grounded) directly into continuous sensory variables such as sonar and infrared readings.

Mobile robots often have to reason and plan their course of action in unknown, non-deterministic, and partially observable environments. Acquiring a model of the environment that reflects the effect of the robot's actions is essential in such cases. The framework of partially observable Markov decision processes (POMDPs) is especially suitable for representing stochastic models of dynamic systems.

Reasoning, planning, and learning in fully-observable state spaces is relatively well understood, and finding optimal policies for problems with huge numbers of states is now feasible. On the contrary, planning and learning in problems with partially-observable state spaces have proven to be very difficult. Finding optimal policies for POMDPs has been shown to be PSPACE-hard and currently solvable problems rarely have more than a hundred states. Learning POMDPs is harder still – even for problems with discrete state spaces and discrete observations, the number of states for the largest problems solved are in the dozens.

The usual approach to learning POMDPs has been to unfold the model in time and employ a general method for learning probabilistic networks such as Baum-Welch (BW), which tunes the parameters of the model so as to increase the likelihood of the training data given these model parameters. Such algorithms often get trapped in shallow local maxima of the optimization surface and can rarely recover the true

POMDP that has generated the data. In an attempt to find a learning method that can overcome this problem, we have adapted the best-first model merging (BFMM) algorithm, originally proposed by Stolcke and Omohundro for learning hidden Markov models. An extensive experimental study over 14 synthetic worlds showed that this algorithm has advantages over the BW for learning POMDPs (Nikovski and Nourbakhsh 1999).

In spite of its advantages, the BFMM algorithm also leads to local maxima in likelihood, because it selects mergers greedily and never reconsiders suboptimal ones. We have developed a new algorithm for learning POMDPs based on merging states by performing clustering in the space of trajectories of percepts and actions leading into and out of hidden states. The advantage of this algorithm is that it considers all possible mergers at once and chooses only the best ones, based on similarities between trajectories. We have compared experimentally this algorithm with BW and BFMM in a simulated environment resembling a typical office space with moderate perceptual aliasing (Nikovski and Nourbakhsh 2000). Seven similarity measures have been tested and one of them yielded results far exceeding those of the other algorithms.

We are presently working on reducing the computational complexity of the learning algorithm from $O(N^3)$ to $O(N \log N)$, where N is the length of the observation/action sequence. We are also working on modifying the algorithm to handle continuous observations by designing similarity measures between pairs of trajectories consisting of continuous sensor readings, using several techniques from the fields of system identification and optimal control.

References

- 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*, 137–143. IEEE, Monterey, CA, 1999.
- D. Nikovski and I. Nourbakhsh. Learning probabilistic models for decision-theoretic navigation of mobile robots. To be published in *Proceedings of the Seventeenth International Conference in Machine Learning*. Stanford, CA, 2000.