Learning Rich, Tractable Models of the Real World
MIT9904-09

Progress Report: January 1, 2001—June 30, 2001

Leslie Pack Kaelbling

Project Overview

The everyday world of a household or a city street is exceedingly complex and dynamic, from a robot's perspective. In order for robots to operate effectively in such domains, they have to learn models of how the world works and use them to predict the effects of their actions. In traditional AI, such models were represented in first-order logic and related languages; they had no representation of the inherent uncertainty in the world and were not connected up to real perceptual systems. More recent AI techniques allow model-learning directly from perceptual data, but they are representationally impoverished, lacking the ability to refer to objects as such, or to make relational generalizations of the form: "If object A is on object B, then if I move object B, object A will probably move too."

We are engaged in developing learning algorithms that can represent and acquire such information from interaction with a noisy real world.

Progress Through June 2001

In September, 2000, we added a postdoctoral researcher, Tim Oates, and two research assistants, Natalia Hernandez and Sarah Finney to the project. We spent the first three months exploring the literature and developing a concrete research agenda.

In January, we began to carry out a set of experiments exploring the use of deictic propositional representations in a simple blocks-world environment. The fundamental idea is that, rather than describing the objects in the world with arbitrary names, such as "block 23", it makes more sense to describe them in terms related to the agent, such as "the block I am looking at". There had been some existing studies of learning with deictic representations, but they were small and the representations seemed to be very carefully hand-tuned to work in the domain. We wanted to explore the use of general reinforcement-learning techniques in these domains, without hand-crafting the representations. We had two goals: comparing a propositional deictic representation to a more traditional propositional representation, and understanding how different reinforcement-learning methods (in particular, NeuroDyanmic programming and Utree) interacted with deictic representations.

We made a great deal of progress during this period, but unfortunately, most of our results were negative. That is, we found that none of the different techniques we tried worked very well in the block-stacking domain. After the initial set of experiments, we did a great many more experiments, aimed at elucidating exactly the causes of failure of the initial experiments. We have learned a great deal through all of this. The details are described in a technical report; but the major points are listed below:

- The statistical tests in the Utree algorithm can be made much simpler and more reliable
- Chapman and Kaelbling's G algorithm can be applied in similar circumstances to the Utree algorithm; it is simpler to implement and more computationally efficient, but it does not use data as efficiently as Utree
- The G algorithm (and Utree) tend to grow much larger trees than necessary, especially when they are allowed to split on observations from preceding time steps. This is a serious handicap to their application.
- The deictic representation is highly partially observable, and requires policies with more steps (because active perception steps are also needed). This makes exploration quite difficult and makes the deictic learning methods almost completely impractical. McCallum's initial experiments with learning in blocks-world with a deictic representation used human guidance in exploration; now we understand why.
- The propositional representation grows as distracter blocks are added to the domain and is also impractical.

Our conclusions from all of this work are that it is important to identify something about the structure of the domain and how the observations work, rather than trying to learn policies directly via reinforcement learning.

Research Plan for the Next Six Months

In the next six months (which will be the final period of this project), our plan is to investigate methods for learning forward models of the dynamics of the domain, rather than trying to learn the reinforcement function directly. If we are able to learn a model, then we can use it to build a state estimator, and then render the problem completely observable. This is a big problem, but we have the following concrete steps planned:

- Starting from deterministic finite-state automaton learning algorithms (such as those of Rivest and Schapire), extend them to:
  - Work in probabilistic domains,
  - Learn factored representations;
  - Be driven by reinforcement.
- Consider appropriate representations of belief state, and methods for updating them, based on the model-learning work described above.
- Prepare and submit a paper on last period's negative results to an international conference
- Study the role of probability in these representations, with particular emphasis on making an efficient near-deterministic approximation, but generalizing robustly when necessary.
- Design vision algorithms that learn to do object segmentation based first on optical flow, with a goal of making them more robust by allowing the agent to manipulate the objects