Partially Observable Markov Decision Processes

Tutorial | Papers | Talks | Code | Repository
Back | POMDP Tutorial | Next

Background on POMDPs

We assume that the reader is familiar with the value iteration algorithm for regular discrete Markov decision processes (MDPs). However, we will need to differentiate these from POMDPs which we could also call a discrete Markov decision process. Therefore, we will refer to the more familiar MDPs as CO-MDPs, emphasizing that they are completely observable.

Adding partial observability to an MDP is not a trivial addition. Solution procedures for CO-MDPs give values or policies for each state. Use of these solutions requires the state to be completely known at all times and with complete observability this presents no problem. Partial observability clouds the idea of the current state. No longer is there certainty about the current state which makes selecting actions based on the current state (as in a CO-MDP) no longer valid.

A POMDP is really just an MDP; we have a set of states, a set of actions, transitions and immediate rewards. The actions' effects on the state in a POMDP is exactly the same as in an MDP. The only difference is in whether or not we can observe the current state of the process. In a POMDP we add a set of observations to the model. So instead of directly observing the current state, the state gives us an observation which provides a hint about what state it is in. The observations can be probabilistic; so we need to also specify an observation function. This observation function simply tells us the probability of each observation for each state in the model. We can also have the observation likelihood depend on the action if we like.

Although the underlying dynamics of the POMDP are still Markovian, since we have no direct access to the current state, our decisions require keeping track of (possibly) the entire history of the process, making this a non-Markovian process. The history at a given point in time is comprised of our knowledge about our starting situation, all actions performed and all observations seen.

Fortunately, it turns out that simply maintaining a probability distribution over all of the states provides us with the same information as if we maintained the complete history. In a CO-MDP we track our current state and update it after each action. Here this is trivial, because it is completely observable. In a POMDP we have to maintain this probability distribution over states. When we perform and action and make an observation, we have to update the distribution. Updating the distribution is very easy and just involves using the transition and observation probabilities. You'll have to take our word for this, since we are prohibited from showing you the formula.

Continue

Back | POMDP Tutorial | Next

Last modified: Thu Nov 6 23:06:44 CST 2003