Partially Observable Markov Decision Processes

Brief Introduction to the Value Iteration Algorithm

With MDPs we have a set of states, a set of actions to choose from, and immediate reward function and a probabilistic state transition function. Our goal is to derive a mapping from states to actions, which represents the best actions to take for each state, for a given horizon length. This is easy if we know the value of each state for this horizon length. The value iteration algorithm computes this value function by finding a sequence of value functions, each one derived from the previous one.

The value iteration algorithm starts by trying to find the value function for a horizon length of 1. This will be the value of each state given that we only need to make a single decision. There isn't much to do to find this in an MDP. Recall that we have the immediate rewards, which specify how good each action is in each state. Since our horizon length is 1, we do not need to consider any future effects (there is no future). Thus, we can simply look at the immediate rewards and choose the action with the highest immediate value for each state.

The next step, which is the second iteration of the algorithm, is to determine the value function for a horizon length of 2. The value of acting when there are two steps to go, is the immediate rewards for the immediate action you will take, plus the value of the next action you choose. Conveniently, we have already computed the values of each state for a horizon length of 1. So to find the value for horizon 2, we can just add the immediate effects of each of the possible actions to the already computed value function to find the action with the best value given that there will be two decision to be made.

Note that we will have to figure in the probabilistic effects of the actions that might happen as we go from having 2 actions left to having 1 action left. Thus, we will really need to do a probabilistic weighting, or expected value computation when using the horizon length 1 solution to help solve the horizon length 2 solution.

Now the algorithm iterates again; it finds the horizon 3 value function using the horizon 2 value function. This iterates until we have found the value function for the desired horizon.

We could give the formula for how to calculate one value function from another, but we promised there would be no formulas. All we will say is that it involves iterating over all the states and using the transition probabilities to weight the values. Anyway, the formulas can be found most anywhere people talk about MDPs, so just be content that it is easy to derive one value function from another by iterating over the states.

Continue

Back | **POMDP Tutorial** | **Next**

Last modified: Thu Nov 6 23:00:58 CST 2003