CS-E4800 AI Exercises 7: Partial Observability February 28, 2017

1 Exercise

Consider the following problem. There are 3 identical-looking objects (call them A, B and C). It is known that two of them have the same weight, but the third is either heavier or lighter.

The states in this problem can be viewed as vectors of three weights (for A, B and C), with 2 denoting the weight of the two-equal weighted objects, and 1 denoting a lighter object and 3 a heavier object.

The only action obtains information about the weigths with a scale, where you can place some object(s) X on the right side and some other(s) Y on the left side, and the scale yields observations $X > Y$, $X = Y$ or $X < Y$.

1. What is the belief state initially?
2. Devise a conditional plan for determining which object is the odd one, and whether it is heavier or lighter than the other two. Indicate what is the belief state at each stage of every execution of your plan. It is probably easiest to depict the plan in the form of a decision tree, where the three results of a weighting lead to different subtrees. Developing the plan becomes easy if you keep track of the belief state on every possible execution path during its construction.

2 Exercise

A patient has entered the hospital, and the most visible symptoms suggest two possible diseases A and B, respectively with probabilities 0.01 and 0.1. We make a simplifying assumption that the patient cannot have both A and B at the same time. With the remaining probability 0.89 the patient has neither of the diseases.

1. A test is ordered, with a result O_1, and with the diseases displaying this result with the following probabilities.

$$\begin{aligned} P(O_1|A) &= 0.9 \\ P(O_1|B) &= 0.05 \\ P(O_1|\text{no disease}) &= 0.0 \end{aligned}$$

   Calculate the new belief state.
2. A second test is ordered, with the result O_2 and the following observation probabilities.

$$\begin{aligned} P(O_2|A) &= 0.02 \\ P(O_2|B) &= 0.95 \\ P(O_2|\text{no disease}) &= 0.0 \end{aligned}$$

   Calculate the belief state after the second test.

3 Exercise

Finding best possible policies limited to some specific narrow class can be much easier than to find arbitrary POMDP policies (so-called *history-dependent* policies, which have to take into account an arbitrarily long sequence of past observations when choosing an actions.)

Policies with *small memory* or *bounded memory* is one such class. The idea is that the space of all belief states is partitioned to a small set of discrete states (e.g. 5 or 10 or 50), and policies are defined as mappings *memory state* × *observation* → *action* × *new memory state*.

A simple special case of bounded memory policies is *memoryless policies*, which are mappings *observation* → *action*, with no memory involved at all.

Consider a POMDP with the following properties.

- Two states s_0 and s_1
- Two actions Go and Stay (like the example in the lecture)

- Transition probabilities are as follows.

$$\begin{aligned} P(s_0, Go, s_1) &= 0.9 \\ P(s_0, Go, s_0) &= 0.1 \\ P(s_1, Go, s_0) &= 1.0 \\ P(s_1, Go, s_1) &= 0.0 \\ P(s_0, Stay, s_0) &= 0.9 \\ P(s_0, Stay, s_1) &= 0.1 \\ P(s_1, Stay, s_1) &= 1.0 \\ P(s_1, Stay, s_0) &= 0.0 \end{aligned}$$

- Reward 2 is obtained when reaching state s_1. No reward for s_0.
- By probability 0.8 observation s_i is made when reaching state s_i, and otherwise observation s_{1-i} is made.

There are 4 different memoryless policies for this problem.

	observation	
policy	s_0	s_1
0	Stay	Stay
1	Stay	Go
2	Go	Stay
3	Go	Go

Each policy induces a *Markov chain*, a stochastic process that generates some state sequence e.g. $s_0, s_1, s_1, s_1, s_0, s_1, \ldots$, that has – for each state – a fixed probability distribution of successor states (i.e. no choice of actions as in MDPs). Essentially, in every state a certain observation is made (stochastically), and the observation uniquely determines the action, which determines a probability distribution over successor states.

1. Evaluate the value of each state under the Markov chains corresponding to policies 1 and 2 (or all four, if you have time and interest to do more work), with discount factor $\gamma = 0.7$. This can be done by solving a system of linear inequations exactly like when we evaluated policies for policy iteration earlier.
2. Depict the value functions of these two policies for all belief states as in the POMDP lecture as a curve (relative probabilities of s_0 and s_1 on x-axis, and the values for every possible belief state on y-axis.)
3. We assume that an initial observation is made based on the actual state, and the actual state follows a probability distribution over s_0 and s_1, the *initial belief state*.
   Which policy to choose? Is the choice dependent on the initial belief state?