Decision Making in Robots and Autonomous Agents (INFR11090): Homework Assignment 2 (Semester 2 - 2017/18)

Subramanian Ramamoorthy

4 March 2018

1 Instructions

- This homework assignment is to be done *individually*, without help from your classmates or others.
- Solve all problems and provide your **complete** solutions (with adequate reasoning behind each step, and citations where needed) in a computer-printed form.
- Submit electronically (including all electronic files associated with the assignment), **as well as** submitting a hard copy of your report to ITO.
- This assignment will count for 20% of your final course mark. It is due at 4 pm on 29 March 2018.

Good Scholarly Practice: Please remember the University requirement as regards all assessed work for credit. Details about this can be found at:

`http://web.inf.ed.ac.uk/infweb/admin/policies/academic-misconduct`

2 Specification

This coursework should be completed using the notebook found at `https://infr11090.github.io/cw`

2.1 Question 1 [15 marks]

You are to design an agent that will be in charge of the navigation of a submarine, through a minefield, towards a goal.

Navigating through uncharted territory can be dangerous, which we will model by applying some damage to the submarine. Additionally, there is some uncertainty in the execution of actions when moving through uncharted territory. Currents are very strong and there is a 20% chance that the submarine will diverge away from its path during execution. For example, when executing a go *down* action (command) within the uncharted territory, there is an 80% chance the submarine will go downwards, but also a 10% chance that it will go left, and another 10% chance it will go right.

Hitting a mine will lead to the destruction of the submarine, and boundary regions marked **edge** or **sky** cannot be traversed. Your task is to develop a strategy for the submarine agent (plotted as a *blue* dot) to safely navigate towards the goal (plotted as a *green* dot).

You are to do this using the principles of Dynamic Programming. In particular, in this exercise, you will implement the methods of Policy and Value Iteration.

For all of these questions, please use the provided visualisation functions to plot your policies and their associated trajectories. Include concise comments within the notebook to highlight any interesting observations to support your answers!

2.1.1 Reward and Transition Functions

Define suitable reward and transition functions for the submarine environment. This is a key task for you as the designer of this agent.

The reward function specifies a scalar reward for every state in the environment (in accordance to the coursework description and environment definition above). The transition function determines the next **feasible** state, s', reached when an action a is taken from state s.

2.1.2 Value Iteration

Your first algorithmic task will be to apply value iteration to the solution of this navigation task, using the reward and transition functions you devised, (in the previous step) within the given specification of the environment.

2.1.3 Policy Evaluation

Define a function that computes the state-value function V^{π} on this environment when given a random policy π.

2.1.4 Policy Iteration

Similar to the earlier question, now apply policy iteration to solving the same navigation task. In what ways are the results of policy iteration different from those of value iteration?

2.1.5 Hyper-parameters

For both policy and value iteration:

1. How does varying the discount factor affect the calculated policy? Repeat the same experiment with at least three different settings of a discount factor in order to make your argument.
2. Given the insights from your experiment above, suggest a suitable strategy for setting an appropriate discount rate?

2.1.6 Noisy trigger

The mines now have a 50% chance of blowing up if you are within their direct vicinity (cells on their border). How does this affect your policy? Implement this and justify your answer based on the results.

2.2 Question 2 [5 marks]

Now consider two additions to the above problem specification, that are common in realistic versions of this problem:

1. The agent must include within the costs, the ides of navigability - describing the fact that some parts of a given map are easier to travel through than others (typically, due to the 'difficulty' of performing low-level control with respect to the terrain features found at that depth). One way to model this would be to include a distribution of costs over the given terrain map. By re-writing your complete problem specification, explain how you will incorporate this feature in your modelling.

2. The agents in the environment, previously viewed as static obstacles defining where you can and can not traverse, could have their own dynamics. For instance, one of the mines could be an active craft with its own motion policy. Assuming that you can observe the *current* position of this active craft through a *noisy* channel (e.g., by interpretation of sonar reflections), pose your - now interactive - motion planning problem in Bayesian terms and explain how your solution strategy might need to be altered.

Remark: This second question need not be implemented in the notebook, in the same way you wrote programs for the first question. Instead, we expect to see a complete problem formulation, with suitable expressions and graphs, and an explanation of the solution procedure in terms of those expressions.