6CCS3AIN, 2019, Tutorial 04 (Version 1.0)

1. I have half an hour to spare in my busy schedule, and I have a choice between working quietly in my office and going out for a coffee.

   If I stay in my office, three things can happen: I can get some work done ($Utility = 8$), or I can get distracted looking at the US mid-term election forecast ($Utility = 1$), or a colleague might stop by to talk about some work we are doing on revising the curriculum ($Utility = 5$).

   If I go out for coffee, I will most likely enjoy a good cup of smooth caffienation ($Utility = 10$), but there is also a chance I will end up spilling coffee all over myself ($Utility = -20$).

   The probability of getting work done if I choose to stay in the office is 0.5, while the probabilities of getting distracted, and a colleague stopping by are 0.3 and 0.2 respectively.

   If I go out for a coffee, my chance of enjoying my beverage is 0.95, and the chance of spilling my drink is 0.05.

   (a) Compute the expected utility of staying in my office and of going out for a coffee.

   (b) By the principle of maximum expected utility, which action should I choose?

   (c) Would this decision change if I use the maximin or maximax decision criteria?

2. Consider the simple world that we studied in the lecture (Figure 1).

   (a) Write down a formal description of this as a Markov Decision Process (as in the slides).

   (b) Assume that actions are deterministic (so the agent moves with probability 1 in the direction it is trying to move) write down a version of the Bellman equation that would work in this case.

   Hint: Take the Bellman equation from the slides and simplify it so that for each a, $P(s'|s, a) = 1$ for one pair of s and s' and 0 for all other pairs.

   (c) Use this deterministic version of Bellman to run value iteration on the world, and obtain utility values for each state.

   You will need to run value iteration until it stabilises.

   (d) Write down the optimum policy given your solution to the deterministic version of value iteration.

3. Now consider the same world, but now assume that actions are non-deterministic. For any action, the action succeeds with probability 0.9 and the action completely fails with probability 0.1 (so that the agent does not move).

   For example, if the action is Up, the agent moves up with probability 0.9, and stays in the same place with probability 0.1.

   (a) Use the non-deterministic version of Bellman to run value iteration on the world, and obtain utility values for each state.

   Note that this question is asking for the values after the values for all states have converged.

   Hint: A spreadsheet is a good way to simplify the calculation.

   (b) Write down the optimum policy. How does this differ from the optimum policy for non-deterministic model that you calculated in Q2? What does this suggest?

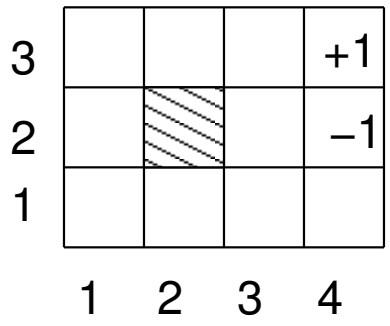

Figure 1: The simple world that we studied in the lecture.

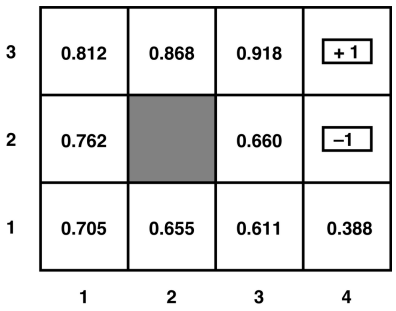

Figure 2: The utility value for the simple world that we studied in the lecture where the problem has the rewards and the transition model from the lecture.

4. Now consider the version of this problem from the lecture — recall that in this version the transition model is such that when the agent tries to move in a given direction, 10% of the time it moves to the left of the chosen direction[1], and 10% of the time it moves to the right. Figure 2 shows the utilities that are obtained under the optimal policy for this version ofthe problem.

   (a) Using the utility values in Figure 2, compute the expected utility of state (3, 1). What do you notice about this value?

   (b) Now compute the action that maximises expected utility.

   (c) Given the utility values of each state in Figure 2, what would the policies be if we chose actions using:

      i. maximin

      ii. maximax

      Compare these to the policy chosen by MEU (this is available in the slides).

5. There is no additional computational part for this tutorial. Partly this is because the spreadsheets in questions 2 and 3 are a form of computation; and partly this is because you will have to implement an MDP solver for the coursework, so you'll get a chance to engage with the computational aspects of Bellman then.

[1]Here "left" and "right" are relative to the direction that the agent is trying to move in. If you find that hard to visualize, you might try thinking of "left" as meaning "90 degress to anticlockwise of" and "right" as "90 degrees to clockwise of".