Project 3
Artificial Intelligence

Fall 2023

[Solutions to this assignment must be submitted via CANVAS prior to midnight on the due date which is 22 November 2023. Submissions up to one day late will be penalized 10% and a further 10% will be applied for the next day late. Submissions will not be accepted if more than two days later than the due date.]

This project may be undertaken in pairs or individually. If working in a pair, state the ***names*** of the two people undertaking the project and the ***contributions*** that each has made. Only ONE submission should be made per group.

Purpose: To gain a thorough understanding of the working of a reinforcement learning agent that assigns a customer for pickup and plots a path to implement the pickup action. In practice any given taxi will in general need to choose which customer to pick up from a pool of requests that are sent to it. Figures 1 provides a view of the environment in which the taxi agent operates.

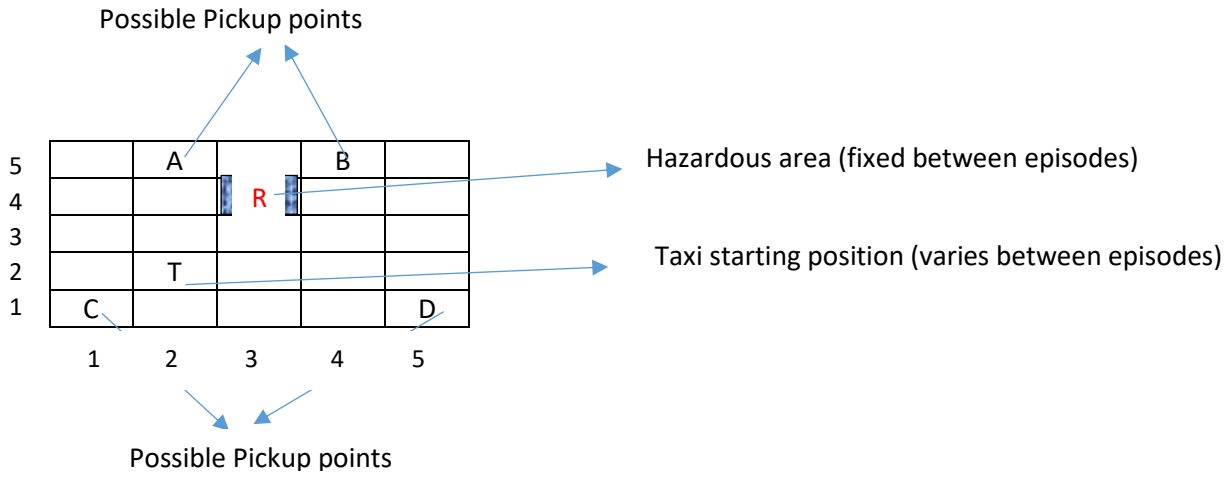

Figure 1

Environment Description: The environment in which the taxi agent operates is a 5 by 5 grid. Passenger pickup points are A, B, C or D as indicated in Figure 1. These positions are fixed across episodes. An episode consists of the agent choosing between customers to pick up (we will assume a maximum of two customers are sent to the taxi agent at any given time) and then navigating from its current location to pick up the chosen passenger. Thus, the taxi starting position may vary between episodes.

Customers are of two different types: regular or premium. A premium customer pays 50% more fare than a regular customer. Thus, if we set the reward for picking up a regular customer at 20, then it is reasonable to set the reward for premium customers at 30. The patterned area R is a restricted area and has a negative reward of -10.

The navigation rule (from any position on the grid) from A, B, C, D and R is as follows. A maximum of four actions, Up, Down, Left and Right are possible from any given position. Navigation in the intended direction occurs with 80% probability and movement in unintended directions (taken as a sum) adding up to 20%. Note that movements never take place opposite (at 180 degrees) to the intended direction. When navigation in the intended direction results in a pickup then the probability rises from 80% to 90% and movement in unintended directions making up 10%. Thus, for example, if the Bellman update was to be applied to state (3,5) (note that indexing is by column first and then by row to maintain

consistency with the lecture notes), then the transition to intended state (4,5) occurs with 90% probability and the agent remains in (3,5) with 10% probability (as the Up action causes the agent to bounce back to (3,5)). The discount factor $\gamma = 0.9$.

Jot down any questions/doubts that you may have and feel free to ask me questions in class or in person. Together with your partner, ***work out a strategy before you start coding the solution in Python.*** Note that this project, unlike the previous two projects, has only one milestone and hence we expect that you on your own will carry on the good practice of designing a solution before implementing it.

Given the limited timeframe for the project, some simplifications have been assumed.

Your task in this project is to implement the following requirements. The project has only ONE milestone (one submission) which has the following requirements.

Requirements

Your Python code in Colab should meet the following requirements. All the outputs required in R1-R4 must be displayed by your Python program. The outputs should also be included in the pdf file.

R1: The agent should pick up a regular customer from position A. In this scenario only one customer has made a request to the agent. The taxi starts in position (2,2). To demonstrate that your code is working, produce the optimal policy. The optimal policy can be visually created in one of two ways: (a) Draw it with arrows showing the path in the same style as discussed in the lectures or if you prefer using (b) Populate each cell in the grid with numeric values produced after running the Value Iteration algorithm.

(20 marks)

R2: The agent should pick up a premium customer from position B. In this scenario only one customer has made a request to the agent. The taxi's start position is (2,2). To demonstrate that your code is working, visualize the optimal policy using your preferred method. **(20 marks)**

R3: The agent should pick up a premium customer from position B. In this scenario two customers have made requests to the agent, a regular customer at point A and a premium customer at position B. The taxi's start position is (2,2). To demonstrate that your code is working, visualize the optimal policy using your preferred method. **(30 marks)**

R4: Run 1000 episodes with random locations of customers. At any given episode, a customer request originating at any of the 4 pickup points is generated at random. A further request (at a point also chosen at random but different from the first chosen point) may happen with probability 0.6. When two requests are received, one of these two (randomly chosen) requests happen to be from a premium customer with probability 0.3. The taxi position should also be chosen randomly from any location in the grid except for position R.

(a) As with requirements R1-R3, your code should display the optimal policy for each episode.
(20 marks)

(b) Compute the fraction of the time that the agent prefers the premium customer over the regular customers over those episodes where both types of customers have made requests for pickup.
(10 marks)

Notes:

Produce ONE pdf document that contains pseudo code, actual Python code and the answers to the questions. Do NOT bury answers to questions as comments in your code. If you are unsure how to produce an original pdf from your Google Co-lab notebook, refer to this tutorial on Youtube: Bing Videos. Do NOT

simply use the print option as this will result in an image. If you submit an image version of a pdf it will not be graded.

End of project specification