MTRE4300 Machine Learning for Robot Perception

Project #8

Due by 11:59 pm on 04/23/21 (Friday)

In this project, you are required to develop a Python program using the reinforcement learning algorithm (Q-learning) to let a robot learn how to navigate in a maze.

1 (Start)	2	3	4
5	6	7 (Trap)	8
9	10	11	12 (Goal)
13	14 (Trap)	15	16

In particular, the project requirements are below:

1. In the first line of your Python code, use a comment line to show all group members' names.
2. The 4x4 maze has 16 grids, numbered from #1 to #16.
3. The robot can move left, right, up, or down but not diagonally.
4. If the robot hits the wall, it will return to its current grid.
5. There are two traps in the maze. If the robot moves in one of them, it will receive a negative score (-10).
6. If the robot moves in the goal (Grid #12), it will receive a positive score (+20).
7. If the robot move in other grids (not Grid #12, #7 or #14), it will receive a score of 0.
8. The robot needs to learn how to move from grid #1 (the start position) to grid #12 (Goal) and collect the maximum score.
9. A **probability state transition function** is assumed in this project. For example, if the robot is located in Grid #6 currently and the "move down" action is selected, after the action is executed, the robot has a probability of 80% to move in Grid #10; has a probability of 10% to move in Grid #5, and has a probability of 10% to move in Grid #7.
10. After the robot is trained for 5,000 steps, your Python program prints the path of the robot moving from Grid #1 to Grid #12. Meanwhile, it displays the curve of the history of four Q values (Q(9,1), Q(9,2), Q(9,3) and Q(9,4)).
11. Each group saves your Python code as "Q_Learning_Maze.py" and uploads it to the D2L drop box.

Grading Rubric

- 20 points: The Python code is submitted correctly.
- 30 points: The code runs without any syntax errors.
- 10 points: The probability state transition function is implemented in the Python code.
- 20 point: After 5,000 steps of training, the robot learns a correct path to move from the start position to the goal position and avoid the traps successfully.
- 20 points. The curve of the history of four Q values is displayed. All Q-values converge.