ECS/DSE-427/627: Multi-Agent Reinforcement Learning

Assignment-1

Question 1: (30 marks)

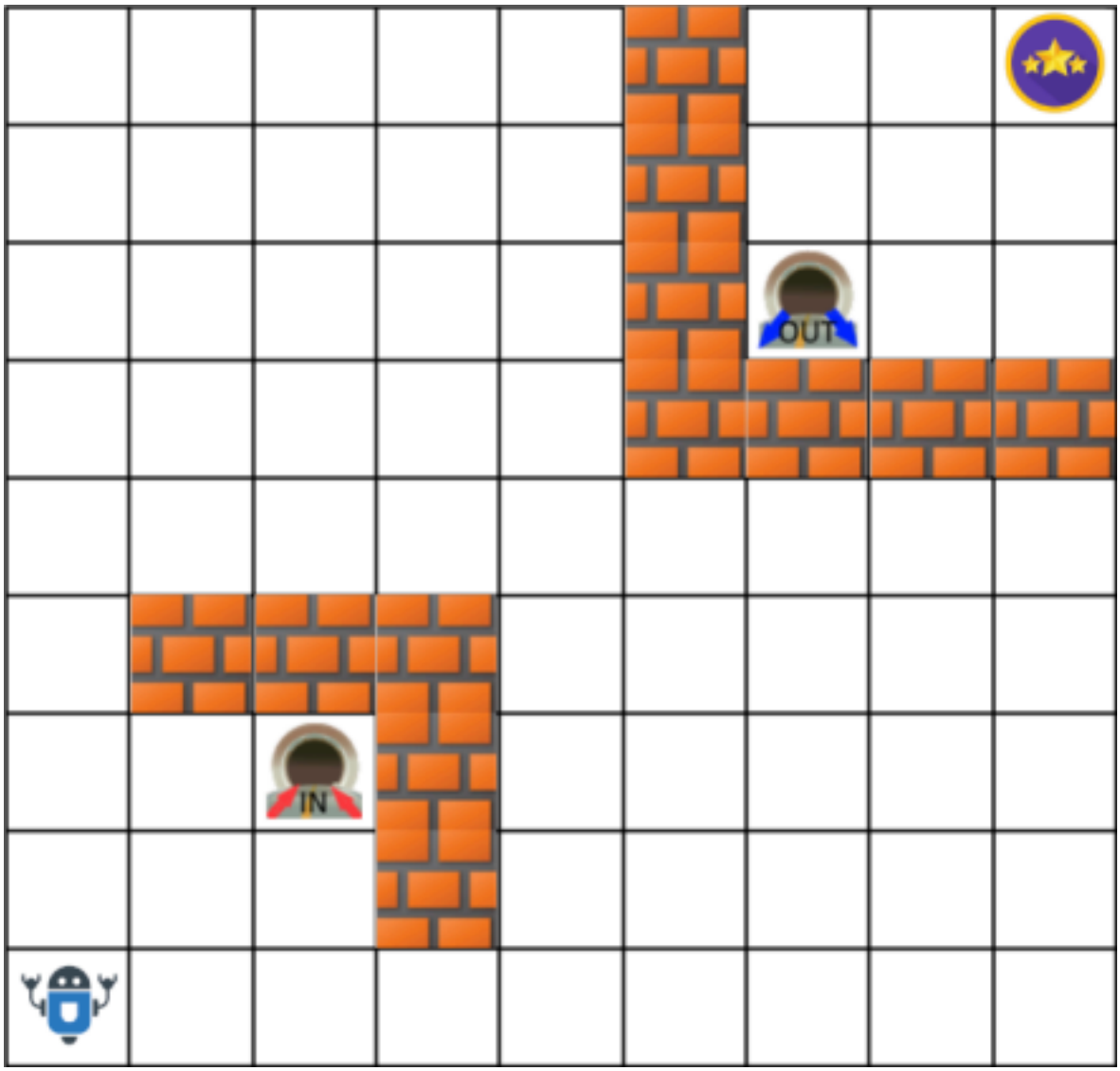
An undergraduate (not-so-ideal) student at the university has the task of attending classes and eating food during his tenure in college. The student has access to three locations on the campus: hostel (reward: -1), academic building (reward: +3), and canteen (reward: +1), and can either eat food or attend class at a given time.

When the student is at the hostel, (s)he attends classes either by going to the academic building with a 50% probability or by staying in the hostel with a 50% probability. When hungry, he/she goes to the canteen (from the hostel) with 100% probability. From the academic building, the student attends class where he stays in the academic building with 70% probability or goes to the canteen with 30% probability. When hungry at the academic building, he/she goes to the canteen with an 80% probability or stays at the same place with a 20% probability. At the canteen, the student has a 60% chance of attending classes by going to the academic building, a 30% chance of attending class by going to the hostel, and a 10% chance of attending from the canteen itself. If hungry, the student will stay in the canteen with a 100% probability.

Using this information, design a finite MDP by writing down the possible combinations of states, actions, transition probability from one state to another for a given action, and rewards in a tabular form. Also, draw a diagram of the MDP from the information mentioning the probability and rewards.
(Refer to example 3.3 of Chapter 3 in Sutton and Barto: Reinforcement Learning)

- Based on the designed MDP, perform value iteration and show the optimal value for each state and the policy obtained.
- Based on the designed MDP, perform policy iteration and show the optimal policy.
- Discuss the results obtained from policy iteration and value iteration.

Question 2: (30 marks)



You are given a 9x9 grid-world environment where:

- The robot icon marks the agent's starting location.
- The star symbol represents the goal position.
- Two tunnels, labelled IN and OUT, serve as one-way portals. The agent can enter through IN and exit through OUT.
- The agent receives a reward of +1 upon reaching the goal; in all other states, the reward is 0.

Your task is to solve this problem using **Value Iteration** and **Policy Iteration** techniques. Specifically, you are required to:

1. Implement both Value Iteration and Policy Iteration to compute the optimal policies.
2. Visualize the optimal policy for each method by plotting a quiver plot, showing the

direction of the agent's optimal movements at each grid cell.