

CSCI 397: Assignment: Implementing and Navigating a Markov Decision Process (MDP)

Due Date: Friday September 29th @ 11:59 PM

Information

In this assignment, you will build a Markov Decision Process (MDP) from scratch and program an agent to navigate through it. Your MDP should have at least 10 different locations, a state transition probability matrix, rewards, and actions. One of the locations should be a terminal state. The agent should traverse the MDP without learning, simply transition from state to state based on the probability and potentially perform an action at that location. You will then run 10 episodes and record the reward, state, and action pair at each step for up to 25 steps.

The scenario is this, you are digging for buried treasure. At each location, there is the potential to dig to find treasure. Your state should include the name of the location and whether treasure exists at the location. You have a 10% chance to dig at any location and if you dig at a location with treasure, you should reward the agent. In the entire environment, there can only be three buried treasures in the environment and their location should not change between episodes. Once the agent reaches a terminal state, the simulation is done, and the cumulative reward should be displayed. Every action should be based on probability rather than learning to accumulate reward.

The actions your agent can take are dig and move. The rewards should be +2 for finding buried treasure, -1 for moving from location to location and +5 for landing on the terminal state. If you land on the terminal state with all three buried treasures, you get a reward of +15. You should introduce gamma within this MDP as part of your reward calculation.

Task 1: Define the MDP

1. Define a set of at least 10 states. One of these should be a terminal state and one of these should be an initial state.
2. Define the state information at each state, which includes the buried treasure locations.
3. Define a set of actions that your agent can take (move, dig).
4. Define a gamma that you will apply at time step.
5. Create a state transition probability matrix to represent the probabilities of transitioning from one state to another.
6. Define a reward function to assign rewards for each state.

Task 2: Implement the Agent

1. Create an agent that starts at an initial state.
2. Implement a fixed policy for the agent based on probability.
3. At each step, your agent should:
 - a. Select whether to act (move states) or not.
 - b. Transition to that new state.
 - c. Record the reward, current state, and action.
 - d. Stop if a terminal state is reached.

Task 3: Run Episodes

1. Run your agent through the MDP for 10 episodes.
2. In each episode, limit the agent to 25 steps or until a terminal state is reached, whichever comes first.
3. Record and print out the history of each episode, including the reward, state, and action at each step.
4. Calculate and print the cumulative reward for each episode.