CSE 546

Reinforcement Learning

Assignment 1 | CheckPoint

Author: Sarveshwar Singhal(50418642)

Date of Submission: Feb 20, 2022

Introduction:

In this project I’ve defined two environments deterministic and stochastic.

Common parameters in both environment:

1. No. of timestamps = 16
2. Env Size: (4\*4) grid. There were in total 16 states defined as follows: {(0,0),(1,0),(2,0),(3,0),(0,1),(1,1),(2,1),(3,1),(0,2),(1,2),(2,2),(3,2),(0,3),(1,3),(2,3),(3,3)}
3. I’ve defined four set of actions: S: {Down, Up, Right, Left}
4. Following rewards were defined: {0, 0.5, 1, 1.5, 2, 2.5, 3}

**Deterministic environment**: These are those type of environments in which if an agent takes a step/move, the agent will definitely end up in that state.

The main objective of defining the deterministic environment is to train our agent in the ideal scenario. Generally deterministic environments are less complex as compared to a stochastic environment. When training our agent, agent training takes less time/computation in a deterministic environment as compared to stochastic (keeping the environment outline same)

If an agent is at state s, and takes an action a, if these two are valid combinations then agent position will be updated otherwise agent will remain at the same position.

If the agent is at state s and takes action a and reaches current position (defined below in table) then the agent will get the corresponding reward.

| Previous state | Current Position | Reward |
| --- | --- | --- |
| Any\* | (0,0) | 0 |
| Any\* | (1,0) | 0.5 |
| Any\* | (2,0) | 1 |
| Any\* | (3,0) | 1.5 |
| Any\* | (0,1) | 0.5 |
| Any\* | (1,1) | 1 |
| Any\* | (2,1) | 1.5 |
| Any\* | (3,1) | 2 |
| Any\* | (0,2) | 1 |
| Any\* | (1,2) | 1.5 |
| Any\* | (2,2) | 2 |
| Any\* | (3,2) | 2.5 |
| Any\* | (0,3) | 1.5 |
| Any\* | (1,3) | 2 |
| Any\* | (2,3) | 2.5 |
| Any\* | (3,3) | 3 |

Any\* describes the state s (from all possible 16 states)

**Stochastic Environment**: These are those type of environments in which if an agent takes a step/move, the agent may or may not end up in the desired state, the environment may move the agent in a different state as well.

The main objective of this type of environment is to train our agent on real world scenarios.

The Stochasticity in this environment is: if agent is at position (0,0), (1,1) … (i.i) then environment generates random probability (for these position only), if the probability is less than 10% then the agent will remain at the same position, otherwise the environment will allow the agent to move normally.

If the agent is at state s and takes action a and reaches current position (defined below in table) then the agent will get the corresponding reward.

| Previous state | Current Position | Reward |
| --- | --- | --- |
| Any\* | (0,0) | 0 |
| Any\* | (1,0) | 0.5 |
| Any\* | (2,0) | 1 |
| Any\* | (3,0) | 1.5 |
| Any\* | (0,1) | 0.5 |
| Any\* | (1,1) | 1 |
| Any\* | (2,1) | 1.5 |
| Any\* | (3,1) | 2 |
| Any\* | (0,2) | 1 |
| Any\* | (1,2) | 1.5 |
| Any\* | (2,2) | 2 |
| Any\* | (3,2) | 2.5 |
| Any\* | (0,3) | 1.5 |
| Any\* | (1,3) | 2 |
| Any\* | (2,3) | 2.5 |
| Any\* | (3,3) | 3 |

Any\* describes the state s (from all possible 16 states)

**Visualization of Environment**:

Deterministic Environment



Stochastic Environment:



**Stochastic environment** was defined based on the randomness in few states of the environment.

The Stochasticity in this environment is: if agent is at position (0,0), (1,1) … (i.i) then environment generates random probability (for these position only), if the probability is less than 10% then the agent will remain at the same position, otherwise the environment will allow the agent to move normally.

**Safety in AI**:

Safety in AI is a major concern, especially when dealing with reinforcement learning agents (because in reinforcement learning we have to deal with uncertainty and new situations). We had to make sure that actions of our agent doesn’t tamper/damages the original environment or it doesn’t create any harm in real world.

In this environment we made sure the agent doesn’t go beyond the specified boundary limits, so by doing this we made sure that our agent is well within our limits and was training in a limited environment, so that it doesn’t harm the external world.