Defining Reinforcement Learning Environment

Ankit Shaw
Compunter Science and Engineering
University at Buffalo
ashaw7@buffalo.edu

Abstract

In Reinforcement Learning (RL), an Environment is the playground where an Agent interacts. Environment have multiple states and the Agent interacts by taking an action to move to the next state. With each action-state pair there is a reward associated which the Agent gets. This reward gives the Agent some hint on whether the action was good or bad. In this paper, we are going to define deterministic and stochastic environment, provide visualization for our custom environments and explain how deterministic environments are different from stochastic environments. Towards the end we will also see how safety checks where put in place for the environments so that the agent only takes actions that are allowed and navigates within the defined state-space. The code for the environment can be found on github.[1]

1 Deterministic Environment

An Environment is said to be Deterministic when next state outcome is determined by the current state and action taken. For instance, in a grid environment if we taken an action left then the agent will always go towards left.

1.1 Definition

1.1.1 States

In our Grid Environment defined the size of the environment is (5 x 5). There are 25 states in total.

$$\text{Set of States: } S = \{S0, S1, S2, \ldots S25\}$$

Each individual cell in the grid is an unique states. These states can have different reward associated to them as well.

1.1.2 Actions

An agent can take four actions in the environment. The actions could be Left, Right, Up, Down. These are denoted by 'L', 'R', 'U', 'D' respectively.

$$\text{Set of Actions: } S = \{\text{'L', 'R', 'U', 'D'}\}$$

1.1.3 Reward

We have five rewards in our grid environment. Two penalty rewards are -3,2 and there are three positive rewards 2, 4, 10.

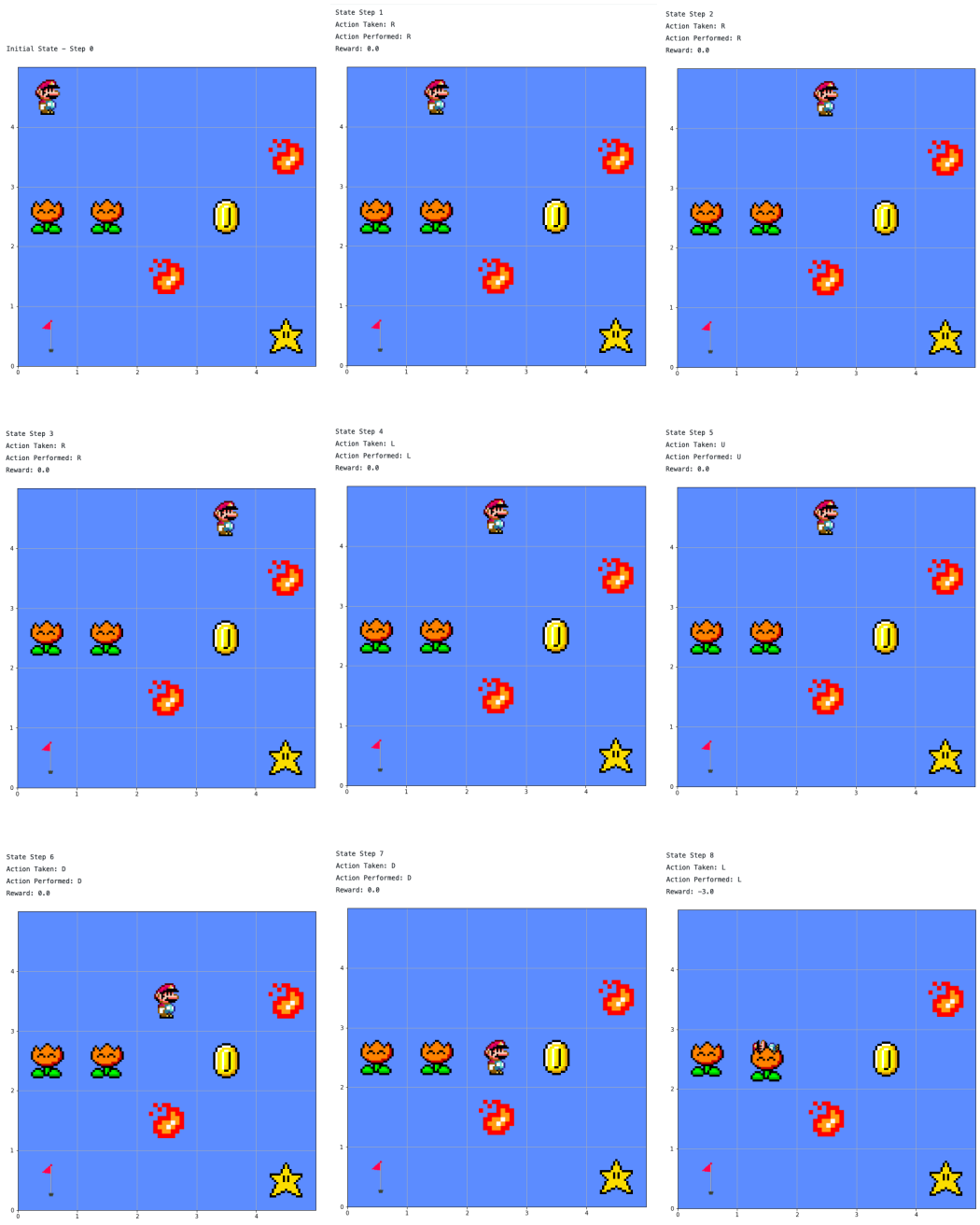
[1] https://github.com/ankitshaw/ub-cse546-reinforcement-learning-a1.git

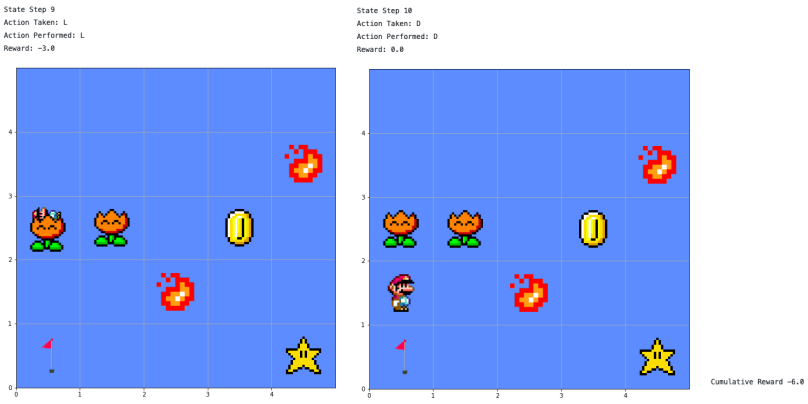
Set of Rewards: R = {-3, -2, 2, 4, 10}

1.2 Objective

The main objective of an agent in the environment is to reach the terminal or goal state and attain the maximum discounted cumulative reward.

1.3 Visualization





2 Stochastic Environment

An Environment is said to be Stochastic when next state outcome cannot not be determined by the current state and action taken. So there is always uncertainty with actions taken. For instance, in a grid environment if we taken an action left then the agent may or may not always go towards left. There is always a probability that the agent may go right or up but not left.

2.1 Definition

2.1.1 States

In our Grid Environment defined the size of the environment is (5 x 5). There are 25 states in total.

$$\text{Set of States: } S = \{S0, S1, S2, \ldots S25\}$$

Each individual cell in the grid is an unique states. These states can have different reward associated to them as well.

2.1.2 Actions

An agent can take four actions in the environment. The actions could be Left, Right, Up, Down. These are denoted by 'L', 'R', 'U', 'D' respectively.

$$\text{Set of Actions: } S = \{\text{'L', 'R', 'U', 'D'}\}$$

2.1.3 Reward

We have five rewards in our grid environment. Two penalty rewards are -3,2 and there are three positive rewards 2, 4, 10.

$$\text{Set of Rewards: } R = \{-3, -2, 2, 4, 10\}$$

2.2 Objective

The main objective of an agent in the stochastic environment is to reach the terminal or goal state and attain the maximum discounted cumulative reward.

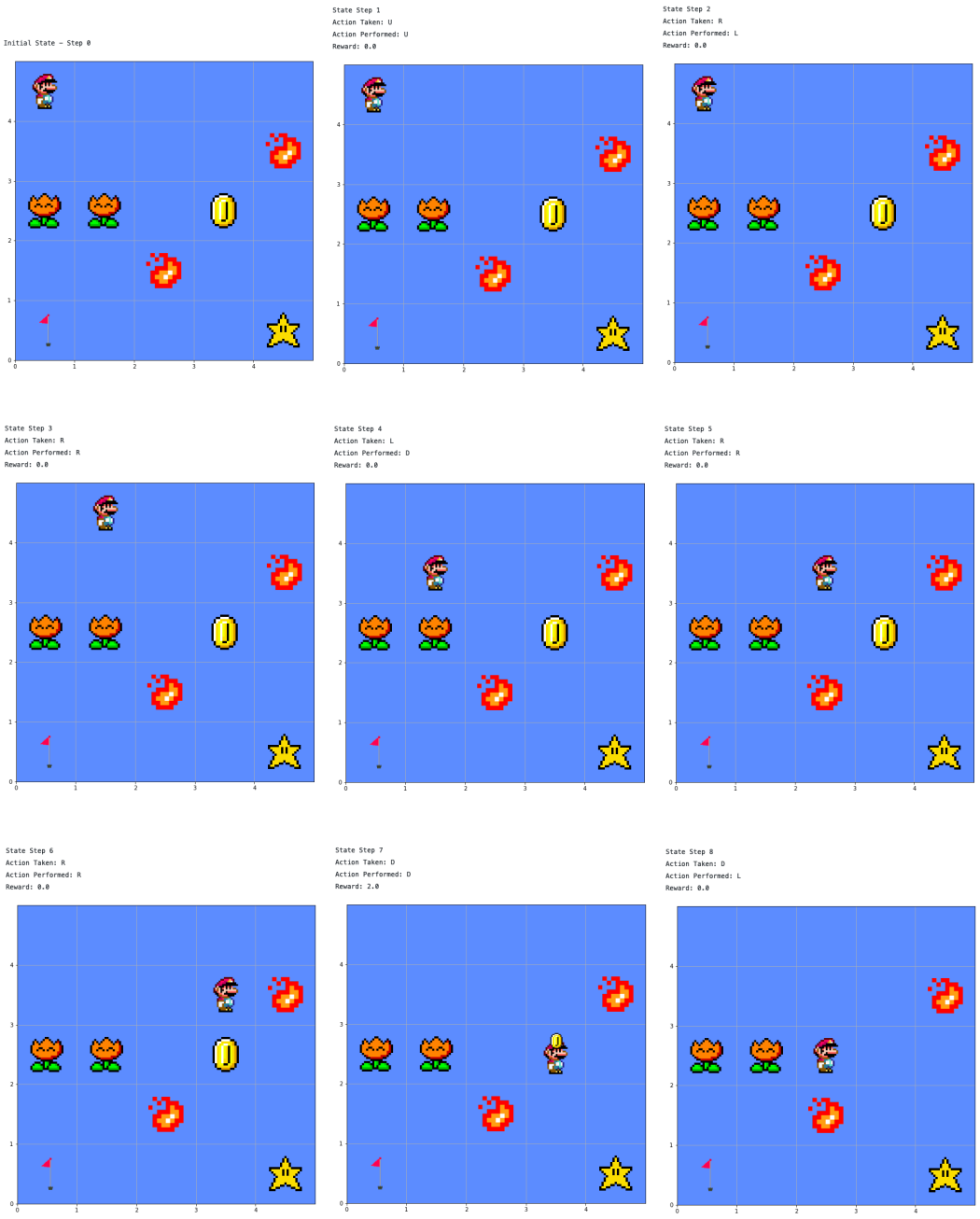
2.3 Stochastic Condition

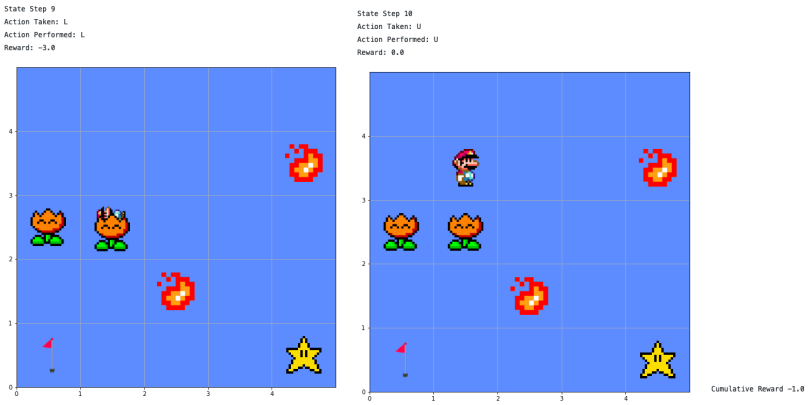
To define the stochastic environment we introduce a parameter: p_transition. p_transition defines the transition probability of taking an action a1 at a given state S1 to move to state s2. To keep things simple we use the same p_transition for all the state-action pair. In simple terms, p_transition tells the probability of executing an action if the action is taken by the agent.

For example: Given a p_transition p, When an agent performs an action a1, then probability that the same action is executed is p. And, the probability executing an action other than a1 is (1 - p_transition)/(action_count - 1).

This is how we introduce the stochasticity or uncertainty in our environment.

2.4 Visualization





3 Difference between Deterministic and Stochastic Environment

The main difference between Deterministic and Stochastic environment is the element of uncertainty associated with an action and state transition. In a Deterministic environment next state outcome is determined by the current state and action taken. For instance, in a grid environment if we taken an action left then the agent will always go towards left. There is no uncertainty in that. While on-the-other in a Stochastic environment next state outcome cannot not be determined by the current state and action taken. There is always an uncertainty associated to the transition. For instance, in a grid environment if we taken an action left then the agent may or may not always go towards left. There is always a probability that the agent may go right or up but not left.

4 Safety in AI

To ensure safety in our environment we design it in such a way that agent interact within a limit of our definition. To ensure that the agent navigates within our defined state-space we do a check on every new state transition to ensure that it lies within the state-space. For our grid environment we check whether the new state coordinates are within the grid shape we defined. Similarly, we filter any actions that are not part of our action set before we execute that action. So if an agent taken an action outside of our action set then the agent remains within that set.

5 Reference

1. Richard S. Sutton and Andrew G. Barto. Reinforcement Learning: An Introduction.
2. Lecture Slides
3. Icons for Rendering credit - Sandro Pereira/ph03nyX, ph03nyx.deviantart.com