Examples Of Markov Decision Process

1. Bubble Shooter Game:

This is an Android/IOS app in which the player has to strike his striking ball of a certain color towards an array of colored balls. If the striking ball’s color matches that of the point where it strikes the array the player scores a positive point and if they don’t match the player receives a negative point.

This process can be visualized as an Episodic Markov Decision Process (MDP) with the following variables. Each episode is defined with starting array and a striker as the initial state and ends when the array is either cleared or reaches the end of the screen.

Action: striking the ball towards the array which is a constant action

States: the color sequence of the array and the color of the present striking ball

Reward: +1 if the color of striker and target match

-1 if the colors don’t match

End of episode: When the array is cleared (game won) or

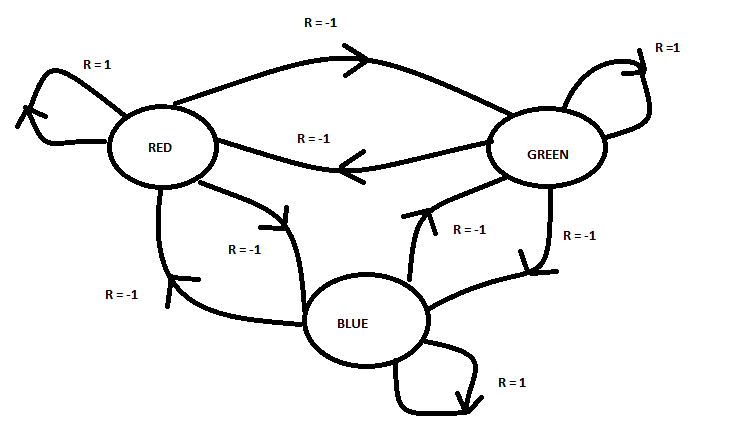
filled until the end of screen (game lost)

Let Blue, Green and Red be the three colors in the array.

The transition Probabilities can be represented as follows:

The probability that the striker gets the same color is alpha = 0.6 and different color is

(1-alpha) =0.4



The process can be shown in a table as follows:

Here previous\_state represents the color of the previous strike and present\_state represents the color of the present strike

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Previous\_state | | Present\_state | | Reward |
| Blue |  | Green |  | -1 |
| Green | Green | 1 |
| Green | Red | -1 |
| Red | Blue | -1 |
| Blue | Blue | 1 |
| **Autonomous UAV Flight :** Consider a scenario where a UAV is to learn to decide its own path from source to destination and in the intermediate path it should drop a package at various checkpoints. Depending on its battery level it can choose to charge or take a flight.  This example can be viewed as an episodic MDP problem. The state, action and rewards can be explained as follows: State : the checkpoint where it previously dropped the package and the battery value(either high or low) Action : the action is to fly autonomously and complete the task Reward : +1 for completing the mission successfully                 0 for charging                  -1 for not completing the mission End of episode : One autonomous flight either successful or not is an episode  Let us consider two checkpoints. If the UAV dropped a packet at Checkpoint 1 (CP = 1) previously now it should drop at checkpoint 2 (CP = 2). The MDP can be expressed as follows :   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | | Checkpoint | | Battery before action | | Action | Reward | Battery after action | | 1 |  | Low |  | Charge | 0 | High | | 2 |  | High | Fly | 1/-1 | Low/High | | 1 |  | Low | Charge | 0 | High | | 1 |  | High | Fly | 1/-1 | Low/High | | 2 |  | High | Fly | 1/-1 | Low/High | |  |  |  |  | |  |  |  |   **Defence Surveilence** Consider a system built to keep track of any enemy attack in a defence military base using RADAR sensors and the system should intelligently take necessary actions.  This task can be visualized as a continuous MDP as the tracking needs to happen continuously. The MDP can be explained as follow: State : Battery Level of the power source Action : Turn on the power supply when battery is low and keep tracking the enemies Reward : +1 for tracking an enemy correctly                 0 for turning on power supply                 -1 for informing a wrong result or not informing an attack  The MDP can be explained as : |  |  |  |  |