Practice Problem Set 1 (Markov Decision Process)

October 6, 2024

1 Batch Transportation (Difficulty level: Easy)

We consider that time is divided into slots. At any given time slot, a transportation company recieves a product with probability p and does not recieve a product with probability $1-p$. After recieving the order, the company stores the product in a warehouse that has a maximum capacity of N products. At any given time slot, the company can decide to book a truck and transport **all** the stored product or not to transport any product at all. It costs the company K INR to book a truck and then an additional cost of c INR per product. The aim of the company is to minimize the expected β-discounted cost over an infinite horizon.

Write the Bellman optimality equation of this problem. Also, give a qualitative explanation of the value function associated with this problem.

2 Rescue Operation in GridWorld (Difficulty level: Medium)

Figure 1: Gridworld showing a rescue boat (grid 20) and a floating survivor wearing life jacket (grid 5).

Consider the gridworld shown in Fig. 1. This gridworld depicts a particular portion of a sea. The grid numbers are shown in the upper-right hand side of the grids. *Please Note: The number of grids does not really matter. The figure is just for your visual reference.*

We consider that time is divided into slots. At a given time slot, a rescue boat that is patrolling the sea recieves a distress signal from a survivor who is floating in a life jacket. The survivor does not know swimming and hence drifting in the sea. Let $p_s\left(x^+|x\right)$ denote the probability of the survivor being in grid number x^+ in the next time slot provided that its grid number is x in the current time slot. We assume that the rescue boat gets continuous update about its location in the sea as well as the location of the survivor.

The rescue boat has to decide whether to go up, down, left, or right in order to minimize the expected search time. One proxy to minimize the expected search time is to minimie the β-discounted cost

$$E\left[\sum_{t=0}^{T} \beta^t c_t\right] \tag{1}$$

where $c_t = 1$ if the grid of the survivor and the boat is not same and $c_t = 0$ otherwise. In (1), T is the first time slot when the grid of the survivor and the boat is the same. Finally, the motion of the rescue boat following an action

$a \in \{up, down, left, right\}$ is captured by the probability distribution $p_b(y^+|y, a)$ which is the probability that the boat will be in grid number y^+ in the next time slot given that its current grid number and current action are y and a respectively. Answer the following questions:

(a) What is the state and state space?

(b) What is the Bellman optimality equation for this problem?

(c) Write the equation for value iteration in order to solve the Bellman optimality equation.

3 Mars Rover Path Planning (Difficulty level: Medium)

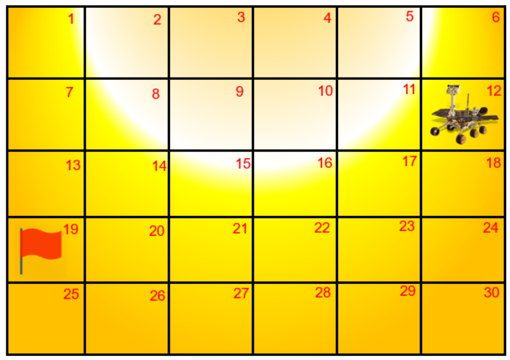

Figure 2: Gridworld showing mars rover (grid 12) and its target grid (grid 19). The shading of the grids shows the level of sunlight the grid recieves.

This problem deals with path planning for mars rover. For most practical scenarios, this problem can be visualized as navigating in a gridworld as shown in Fig. 2. The mars rover has to go from its current position to a target destination as quickly as possible in order to conduct some experiment. In general, the quickest path is a straight line path. However, the twist is that the mars rover is powered solely using solar energy and different grids receive different level of sunlight as shown in Fig. 2. Hence, the mars rover may have to consider a longer path in order a reach the destination quickly! *Please Note: The number of grids does not really matter. The figure is just for your visual reference.*

We consider that time is divided into slots. The mars rover has to reach a target grid $\overline{x}$ starting from its current grid x_0. The mars rover is equipped with a battery of infinite capacity. Let b_t denote the units of battery power left at time slot t. At any given time slot, the battery power of the rover increases by an amount δ due to solar energy. δ is a random variable with probability distribution $p_s(\delta|x)$ where x is the grid location of the rover in that time slot. We assume that the initial battery power and δ are both integers.

At any given time slot, the rover can decide the following:

1. To not move. If the rover does not move, it does not consume any battery power.
2. To move. This means that the rover has to decide the direction and the speed. The direction can be up, down, left, or right. The speed can be high or low. If the speed is low, the rover moves by one grid in the choosen direction and consumes a battery power of θ_l units. If the speed is high, the rover moves by two grids in the choosen direction and consumes a battery power of θ_h units where $\theta_h > \theta_l$. Off course, the rover cannot consume more battery power than it currenly stores.

The objective is to minimize the β-discounted cost given by (1) where $c_t = 1$ if the rover's grid at time t is not equal to $\overline{x}$ and $c_t = 0$ otherwise. Answer the following questions:

(a) What is the state and state space?

(b) What is the action and action space?

(c) What is the Bellman optimality equation for this problem?

4 Previous Year MDP Problems

Go to folder *"previous year exams"* in Dropbox.

(a) In the sub-folder *"minor2"*, open *mock_paper_minor2.pdf*. Solve Q2.

(b) In the sub-folder *"minor2"*, open *minor2.pdf*. Solve Q2.

(c) In the sub-folder *"end_sem"*, open *mock_paper_end_sem.pdf*. Solve Q2.

(d) In the sub-folder *"end_sem"*, open *end_sem.pdf*. Solve Q2.