TÉCNICO LISBOA

Homework 3 - Group 03

Planning, Learning and Intelligent Decision Making

Duarte Almeida 95565

Martim Santos 95638

Exercise 1.

(a) Let Π be the set of possible stops that the truck can visit, Γ set of possible unordered combinations of stops where the trash can appear and Φ the set of possible truck status. We then have:

$$\Pi = \{RP, A, B, C, D, E, F\}, \quad \Gamma = \{\{\}, B, C, D, BC, BD, CD, BCD\} \quad \text{and} \quad \Phi = \{Empty, Full\}$$

We can define the state space $\mathcal{X}$ for the POMDP using the Cartesian product between these three sets as follows:

$$\begin{aligned}\mathcal{X} = \Pi \times \Gamma \times \Phi = \{&(RP, \{\}, Empty),\ (RP, \{\}, Full),\ (RP, B, Empty),\ (RP, B, Full),\\ &(RP, C, Empty),\ (RP, C, Full), (RP, D, Empty),\ (RP, D, Full), \cdots\\ &\cdots (F, CD, Empty),\ (F, CD, Full), (F, BCD, Empty),\ (F, BCD, Full)\}\end{aligned}$$

where $(l, \mathcal{L}, s) \mid l \in \Pi, \mathcal{L} \in \Gamma, p \in \Phi$ represents the state when truck is on stop l, the garbage appeared on stops $\mathcal{L}$ and the truck is in status s.
The action space $\mathcal{A}$ for the POMDP is defined as follows:

$$\boldsymbol{\mathcal{A}} = \{Collect,\ Drop,\ U(p),\ D(own),\ L(eft),\ R(ight)\}$$

And the observation space $\mathcal{Z}$ is defined as:

$$\begin{aligned}\boldsymbol{\mathcal{Z}} = \{&(RP, E), (RP, F), (A, E), (A, F), (B, E, \emptyset), (B, E, G), (B, F, \emptyset), (B, F, G),\\ &(C, E, \emptyset), (C, E, G), (C, F, \emptyset), (C, F, G), (D, E, \emptyset), (D, E, G), (D, F, \emptyset), (D, F, G),\\ &(E, E), (E, F), (F, E), (F, G)\}\end{aligned}$$

where the pairs $(l^{'}, s^{'})$, with $l^{'} \in \Pi,\ s^{'} \in \Phi$ correspond to the location of the truck and its current state (E(empty) or F(ull)). When in locations B, C or D, there is an additional third argument that represents the existence of uncollected garbage in that location (G if there is garbage and $\emptyset$ if there is not).

(b) Let $\mathcal{L}$ denote the set of locations with uncollected garbage – i.e. locations referenced in the first subscript of each state – and let l and s denote the current location of the truck and its loaded state (i.e., if it's empty or full). For every 30 minutes that the garbage remains uncollected in one of the possible locations there is a cost of 0.1 and other periods of time correspond to proportional costs. Therefore, **for a valid action** $a \in \mathcal{A}$ in state $x \in \mathcal{X}$ with uncollected garbage in locations in $\mathcal{L}$, its cost is defined as:

$$c(x, a) = c((l, \mathcal{L}, s), a) = |\mathcal{L}| \cdot \left(\frac{0.1}{30} \cdot t(l, a)\right)$$

where $|\mathcal{L}|$ denotes the number of locations that have uncollected garbage and $t(l, a)$ denotes the time in minutes it takes to execute in l, **given that a is a valid action in l**. For convenience, we establish that $\boldsymbol{t(RP, Collect) = 0}$ and $t(l, a)$ is equal to an arbitrary constant -1 for invalid

actions in that state (which won't make any diference, as we will explain later on). We can thus represent the function t in a matrix $\mathcal{T}$:

$$\mathcal{T} = \begin{array}{c} \\ RP \\ A \\ B \\ C \\ D \\ E \\ F \end{array} \begin{array}{c} \begin{array}{cccccc} Collect & Drop & U & D & L & R \end{array} \\ \begin{bmatrix} -1 & 0 & -1 & -1 & -1 & 30 \\ -1 & -1 & 70 & 55 & 30 & 40 \\ 10 & -1 & -1 & -1 & 40 & 80 \\ 10 & -1 & -1 & -1 & 55 & 55 \\ 10 & -1 & -1 & -1 & 70 & 70 \\ -1 & -1 & -1 & -1 & 55 & 20 \\ -1 & -1 & 70 & 20 & 80 & -1 \end{bmatrix} \end{array}$$

(Note that $c(x, a)$ is at most $3 \cdot \frac{0.1}{30} \cdot 80 = 0.8 < 1$).
If on another hand, **the action is invalid**, the agent incurs in maximum cost (1). Let $i(x, a) = i((l, \mathcal{L}, s), a)$ be a function which is 1 if action a is **invalid** in state x and 0 otherwise. If $\mathcal{L} = \emptyset$ and $s = Empty$, the only valid actions are movements between locations and we can define each value of $i((l, \emptyset, Empty), a)$ in a matrix $\mathcal{I}$:

$$\mathcal{I} = \begin{array}{c} \\ RP \\ A \\ B \\ C \\ D \\ E \\ F \end{array} \begin{array}{c} \begin{array}{cccccc} Collect & Drop & U & D & L & R \end{array} \\ \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

If $s = Full$ and any set $\mathcal{L}$, the validity of the actions is the same as before, but the agent can now drop the trash in RP:

$$i((l, \mathcal{L}, Full), a) = \begin{cases} 0 & \text{if } a = Drop \ \wedge \ l = RP \\ i((l, \emptyset, Empty), a) & \text{otherwise} \end{cases}$$

If $\mathcal{L} \neq \emptyset$ and $s = Empty$, the validity of the actions is the same as in the first case, but the agent can now collect the garbage in any location present in $\mathcal{L}$, i.e.:

$$i((l, \mathcal{L}, Empty), a) = \begin{cases} 0 & \text{if } a = Collect \ \wedge \ l \in \mathcal{L} \\ i((l, \emptyset, Empty), a) & \text{otherwise} \end{cases}$$

Thus, the cost function $c : \mathcal{X} \times \mathcal{A} \to [0, 1]$ admits the following succint representation:

$$c(x, a) = c((l, \mathcal{L}, s), a) = (1 - i(l, a)) \left(|\mathcal{L}| \cdot \left(\frac{0.1}{30} \cdot t(l, a) \right) \right) + i(l, a)$$

(c) For a given hidden state x_t consider an alternative representation $\mathrm{x}_t = (\mathrm{l}_t, \mathrm{s}_t, \mathrm{b}_t, \mathrm{c}_t, \mathrm{d}_t)$, where l_t represents the location of the truck, s_t is random variable that represents the state of truck regarding whether it is full or empty, and b_t, c_t and d_t are binary random variables which takes the value 1 if there is trash at locations B, C and D at time step t and 0 otherwise, respectively.
Similarly, for a observed state z_t consider an alternative representation $\mathrm{z}_t = (\mathrm{l}'_t, \mathrm{s}'_t, \mathrm{t}'_t)$, where l'_t represents the location of the truck driver, s'_t is a random variable that represents the state of truck regarding whether it is full or empty, and t_t is a binary random variables which take the value 1 if there is trash at locations in which the trucker is at time step t (for simplicity of notation, we write the letters of each location/truck state for the values of the corresponding random variables when

we are actually referring to their corresponding numerical encodings).
We are asked to find the belief at time step $t+1$ regarding whether there is garbage in location D. Let $\mathrm{h}_t = \{\mathrm{z}_0, \mathrm{a}_0, \ldots, \mathrm{z}_{t-1}, \mathrm{a}_{t-1}, \mathrm{z}_{t-1}\}$ denote the history of the process up to time step t. Thus, we desire to obtain $b_{t+1}(\mathrm{d}_{t+1} = 1) = \mathbb{P}(\mathrm{d}_{t+1} = 1 \mid \mathrm{h}_{t+1} = h_{t+1})$.
Consider first the case where **the trucker sees that there is garbage to collect in stop D** at time step $t+1$. In that case, $\mathrm{z}_{t+1} = (\mathrm{l}_t^{'}, \mathrm{s}_t^{'}, \mathrm{t}_t^{'}) = (D, s, 1)$ (s can be either *Empty* or *Full*). We thus have that:

$$\begin{aligned}\mathbb{P}(\mathrm{d}_{t+1} = 1 \mid \mathrm{h}_{t+1} = h_{t+1}) &= \mathbb{P}(\mathrm{d}_{t+1} = 1 \mid \mathrm{z}_{t+1} = (D, s, 1),\ \mathrm{a}_t = U,\ \mathrm{h}_t = h_t) \\ &= \mathbb{P}(\mathrm{d}_{t+1} = 1 \mid \mathrm{l}_{t+1}^{'} = D,\ \mathrm{s}_{t+1}^{'} = s,\ \mathrm{t}_{t+1}^{'} = 1,\ \mathrm{a}_t\ = U,\ \mathrm{h}_t = h_t)\end{aligned}$$

Now, if the truck goes to a certain location, it knows for sure the existence of garbage there. Given that the truck is in stop D and observes trash in that location (i.e., $\mathrm{l}_{t+1}^{'} = D$ and $\mathrm{t}_{t+1}^{'} = 1$) it is certain that there is trash D ($\mathrm{d}_{t+1} = 1$). Hence, we have that $\mathbb{P}(\mathrm{d}_{t+1} = 1 \mid \mathrm{h}_{t+1} = h_{t+1}) = 1$.

Finally, consider the case where the truck sees that **there is not garbage to collect in stop D** at time step $t+1$. Analogously to the previous case, we have that:

$$\begin{aligned}\mathbb{P}(\mathrm{d}_{t+1} = 1 \mid \mathrm{h}_{t+1} = h_{t+1}) &= \mathbb{P}(\mathrm{d}_{t+1} = 1 \mid \mathrm{z}_{t+1} = (D, s, 0),\ \mathrm{a}_t = U,\ \mathrm{h}_t = h_t) \\ &= \mathbb{P}(\mathrm{d}_{t+1} = 1 \mid \mathrm{l}_{t+1}^{'} = D,\ \mathrm{s}_{t+1}^{'} = s,\ \mathrm{t}_{t+1}^{'} = 0\ \mathrm{a}_t\ = U,\ \mathrm{h}_t = h_t)\end{aligned}$$

Following the same argument as before, it is impossible that trash exists in stop D at time step $t+1$ given that the truck is in that location in the same time step $t+1$ and does not observe trash, and so the expression above evaluates to 0.

In conclusion, if the **truck observes that there is trash** in stop D at time step $t+1$, it sets its belief in the existence of trash in stop D at time step $t+1$ to **1**; if on another hand the truck **observes that there is no trash** in stop D at time step $t+1$, it sets its belief in the existence of trash in stop D at time step $t+1$ to **0**.