Planning, Learning and Intelligent Decision Making

2022/2023

Homework 2 - Group 25

Guilherme Pereira	105806
Miguel Belbute	96453

Introduction

This project was developed for the course Planning, Learning and Intelligent Decision Making taught at Instituto Superior Técnico under the professor Francisco Saraiva de Melo.

Exercise 1

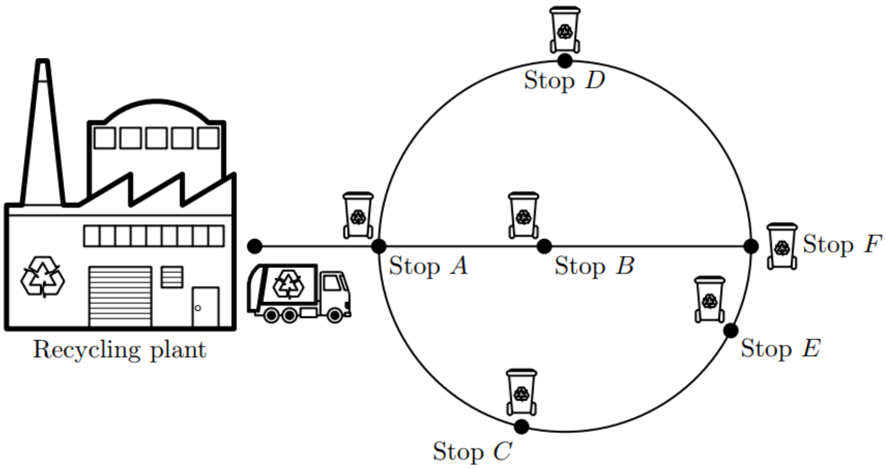

Figure 1: The garbage truck visits the different sites to empty the garbage containers before returning to the recycling plant.

a) The stops and their respective representations are:

- Recycling plant - *RP*
- Stop A - *A*
- Stop B - *B*
- Stop C - *C*
- Stop D - *D*
- Stop E - *E*
- Stop F - *F*

Additionally, *b*, *c*, *d* are a type of flags to indicate if stops *B*, *C* and *D* have already been visited.

Therefore, the **state space** for the MDP is

$$\chi \ = \{(\mathrm{RP}, \sim\mathrm{b}, \sim\mathrm{c}, \sim\mathrm{d}), (\mathrm{RP}, \sim\mathrm{b}, \sim\mathrm{c}, \mathrm{d}), (\mathrm{RP}, \sim\mathrm{b}, \mathrm{c}, \mathrm{d}), (\mathrm{RP}, \mathrm{b}, \mathrm{c}, \mathrm{d}), (A, \sim b, \sim c, \sim d), \ldots (\mathrm{F}, \mathrm{b}, \mathrm{c}, \mathrm{d})\}$$

In other words, it is the combination of all the stops with *b*, *c*, *d*, resulting in a total of *7x2x2x2* = *76* different states.

The available actions and their respective representations are:

- Collect garbage - *CG*
- Drop garbage - *DG*
- Move up - *U*
- Move down - *D*

- Move left - *L*
- Move right - *R*

Therefore, the **action space** for the MDP is

$$A = \{CG, DG, U, D, L, R\}$$

b) Regarding the **cost function** for the MDP, one must consider all the possible actions for each possible state and the following indications:

- The cost associated with an action is proportional to the time that action takes to execute and must be smaller than the maximum cost;
- A successful *DG* has no cost;
- An unsuccessful DG has maximum cost;
- An invalid action (such as *CG* in a stop which has no garbage, like A) has maximum cost;

Assuming a cost function always equal or greater than 0 and less or equal than 1 - ***c(x, a)*** ∊ ***[0, 1]*** -, we can assign a **cost of 0.1 to each 10 minutes taken performing an action**. That being said, we would have the following costs, for example:

+ *c[(B, ~b, c, d), CG] = c[(C, b, ~c, d), CG] = c[(D, b, c, ~d), CG] = 0.1*
+ *c[(E, b, c, d), U] = c[(F, b, c, d), D] = 0.2*
+ *c[(RC, b, c, d), R] = c[(A, b, c, d), L]= 0.3*
+ *c[(B, b, c, d), R] = c[(F, b, c, d), L] = 0.8*
+ *…..*

Unsuccessful or invalid actions would cost 1 (the maximum cost possible):

+ *c[(B, b, c, d), CG] = 1 (collecting garbage once it has already been collected)*
+ *c[(A, b, c, d), CG] = 1 (collecting garbage where there isn't any)*
+ *c[(E, b, c, d), DG] = 1 (dropping garbage in stop different than RP)*
+ *…..*

And a successful action to drop garbage would cost 0 (the minimum cost possible):

+ c[(RP, b, c, d), DG] = 0

Finally, we can try to **summarize this suggested cost function** by writing a branch function. This function, of course, relies on the definition of invalid/unsuccessful and successful actions, which we have previously mentioned and is detailed in the homework assignment.

$$c(\mathrm{x,a}) = \begin{cases} 0, & if\ \mathrm{x}\ =\ (RP, b, c, d) \\ 1, & if\ a\ is\ invalid/unsuccessful \\ t/100, & if\ a\ is\ successful\ and\ t(min)\ is\ time\ taken \end{cases}$$

This definition also depends on the given data - we consider that *t*, the time taken performing an action, is always less than or equal to 80 minutes, which corresponds to the longest possible period of time the agent takes to complete a single action (in this case, traveling between *B* and *F*). With another set of data, there's a possibility we would have to alter the definition… Supposing, for example, the existence of a route which would take over 100 minutes to complete, the cost would go over the current maximum limit of 1.

c) Let us remind ourselves that the cost-to-go function associated with the optimal policy is one such that, for every possible state,

$$J^{\pi^*}_{(x)} \ \leq\ J^{\pi}_{(x)}$$

Moreover, a strictly positive function is one whose values are always greater (and never equal to) 0 - *f(x) > 0 for any x*.

Considering our MDP, we can infer the optimal policy will take into account the agent has to visit stops *B*, *C* and *D*, perform *CB* at each one, return to *RP* and finally perform *DG* with the lowest cost-to-go possible, starting from a known initial state. However, regardless of "trajectory" and sequence of actions, this cost-to-go appears to always be greater than 0. Let's verify this!

Firstly, we can safely state **the optimal policy will never consider invalid/unsuccessful actions** since they drastically increase the cost "paid" by the agent. That being said, **the agent will never act as to remain in the same state** (for example, it will never *CG* in *A*, which is an invalid action, makes the agent pay the maximum value and keeps it in the same state). It is always advantageous to perform valid actions.

If the agent is currently in one of the stops containing garbage (*B*, *C* or *D*), it will either travel to another stop or perform *CG.* None of these actions hold cost 0.

Even if we consider one of the "simplest" initial states possible, *x = (A, b, c, d)* (agent in *A* and with all the garbage collected), the optimal policy would clearly be to reach *RP* and *DG*, but the first action would still hold a cost greater than 0, thusly corresponding to a cost-to-go greater than 0.

However, if we consider the initial state *x = (RP, b, c, d)*, the optimal policy is trivially to just perform *DG*, which holds a null cost and "completes" de MDP! Consequently, in this very specific situation, the **cost-to-go function is equal to 0**, which in turn means it is not strictly positive.

In other words, we can argue that the sentence *“For the MDP above, the cost-to-go function associated with the optimal policy is strictly positive”* is **not true**!

References

[1] - De Melo, F. et al. (2023) Theoretical Lectures, Planning, Learning and Intelligent Decision Making. Instituto Superior Técnico (IST). Available at: https://fenix.tecnico.ulisboa.pt/disciplinas/ADI/2022-2023/2-semestre/material-de-apoio (Accessed: March 4, 2023).