Planeamento, Aprendizagem e Decisão Inteligente

MECD e MMAC

Relatório do Trabalho Prático 2– Parte Teórica

Autores:

Guilherme Lopes (105319) guilherme.n.lopes@tecnico.ulisboa.pt
Leonardo Brito (105257) leonardo.amado.brito@tecnico.ulisboa.pt

Grupo 50

2022/2023 – 2º Semester, P3

Índice

1 Exercício 1

1.1 Alínea (a)

As ações do problema podem ser definidas da seguinte maneira: $\mathcal{A}$ = **{ DG, CG, U, D, L, R }**
Onde,

- DG - Deitar o lixo fora;
- CG - Recolher o lixo;
- U - Ir para cima;
- D - Ir para baixo;
- L - Ir para a esquerda;
- R - Ir para a direita;

Na representação de todos os estados, a letra K representa as localizações e as restantes três letras representam, de forma binária, se o lixo já foi recolhido ou não. O espaço de estados é composto por todas as 56 combinações das 7 localizações e dos três números binários.
.

$\mathcal{X}$ **= { (*K*, d, b, c))}**

Onde,
$K \in$ { A, B, C, D, E, F, RP }

- A - localização A;
- B - localização B;
- C - localização C;
- D - localização D;
- E - localização E;
- F - localização F;
- RP - localização da central de reciclagem (recycling plant);

$$d = \begin{cases} 1, & \text{se o lixo foi recolhido no estado D} \\ 0, & \text{caso contrário} \end{cases} \tag{1}$$

$$b = \begin{cases} 1, & \text{se o lixo foi recolhido no estado B} \\ 0, & \text{caso contrário} \end{cases} \tag{2}$$

$$c = \begin{cases} 1, & \text{se o lixo foi recolhido no estado C} \\ 0, & \text{caso contrário} \end{cases} \tag{3}$$

1.2 Alínea (b)

O custo foi todo transformado em minutos e posteriormente para um número decimal. A função de custo pode ser definida por cf(x, a) $\in$ [0,1] onde, $x \in \mathcal{X}$, $a \in \mathcal{A}$. Quando não usamos a condição *and* nas hipóteses, assumimos elas funcionam para qualquer que seja a combinação dos binários d, b, c.

Quando a localização é A:

$$cf((A, d, b, c), a) = \begin{cases} 0.70, & \text{se } a = \text{U} \\ 0.30, & \text{se } a = \text{L} \\ 0.40, & \text{se } a = \text{R} \\ 0.55, & \text{se } a = \text{D} \\ 1, & \text{caso contrário} \end{cases} \tag{4}$$

Quando a localização é B:

$$cf((B, d, b, c), a) = \begin{cases} 0.40, & \text{se } a = \text{L} \\ 0.80, & \text{se } a = \text{R} \\ 0.10, & \text{se } a = \text{CG} \wedge \text{b} = 0 \\ 1, & \text{caso contrário} \end{cases} \tag{5}$$

Quando a localização é C:

$$cf((C, d, b, c), a) = \begin{cases} 0.55, & \text{se } a = \text{L} \\ 0.55, & \text{se } a = \text{R} \\ 0.10, & \text{se } a = \text{CG} \wedge \text{c} = 0 \\ 1, & \text{caso contrário} \end{cases} \tag{6}$$

Quando a localização é D:

$$cf(D, d, b, c), a) = \begin{cases} 0.70, & \text{se } a = \text{L} \\ 0.70, & \text{se } a = \text{R} \\ 0.10, & \text{se } a = \text{CG} \wedge \text{d} = 0 \\ 1, & \text{caso contrário} \end{cases} \tag{7}$$

Quando a localização é E:

$$cf(E, d, b, c), a) = \begin{cases} 0.55, & \text{se } a = \text{L} \\ 0.20, & \text{se } a = \text{R} \\ 1, & \text{caso contrário} \end{cases} \tag{8}$$

Quando a localização é F:

$$cf(F, d, b, c), a) = \begin{cases} 0.70, & \text{se } a = \text{U} \\ 0.80, & \text{se } a = \text{L} \\ 0.20, & \text{se } a = \text{D} \\ 1, & \text{caso contrário} \end{cases} \tag{9}$$

Quando a localização é RP:

$$cf(RP, d, b, c), a) = \begin{cases} 0.30, & \text{se } a = \text{R} \\ 0, & \text{se } a = \text{DG} \wedge \text{d,b,c} = 1 \\ 1, & \text{caso contrário} \end{cases} \tag{10}$$

1.3 Alínea (c)

Uma política estacionária π^* é ótima se:

$$\forall x \in \chi, \, \forall \pi \in \Pi^S \;\; J^{\pi^*}(x) \leqslant J^{\pi}(x), \;\; \textbf{onde},$$

$$J^{\pi}(x) = \mathbb{E}_{\pi}[\sum_{t=0}^{\infty} c_t \mid x_0 = x] \; = c_{\pi}(x) + \sum_{y \in \chi} P_{\pi}(y \mid x) J(y)$$

Seja $x \in \chi \setminus (RP, 1, 1, 1)$ então:

$$\forall a \in \mathcal{A} \;\; c(x, a) > 0 \Rightarrow \forall \pi \in \Pi^S \;\; c_{\pi}(x) > 0 \Rightarrow \forall \pi \in \Pi^S \;\; J^{\pi}(x) > 0 \Rightarrow J^{\pi^*}(x) > 0 \;\; (1)$$

Suponhamos que escolhemos uma política π^1 com $\pi^1((RP, 1, 1, 1), DG) < 1$, então:

$$\pi^1((RP, 1, 1, 1), DG) < 1 \Rightarrow c_{\pi^1}((RP, 1, 1, 1)) > 0 \Rightarrow J^{\pi^1}((RP, 1, 1, 1)) > 0 \; (2)$$

Suponhamos, agora, que escolhemos uma política π^2 tal que $\pi^2((RP, 1, 1, 1), DG) = 1$, então:

$$\pi^2((RP, 1, 1, 1), DG) = 1 \Rightarrow c_{\pi^2}((RP, 1, 1, 1)) = 0$$

por outro lado sabemos que o sistema reinicia após depositar o lixo:

$$J^{\pi^2}((RP, 1, 1, 1)) = c_{\pi^2}((RP, 1, 1, 1)) + J^{\pi^2}((RP, 0, 0, 0)) = 0 + J^{\pi^2}((RP, 0, 0, 0)) = J^{\pi^2}((RP, 0, 0, 0))$$

mas tendo em conta (1)

$$J^{\pi^2}((RP, 0, 0, 0)) > 0 \Rightarrow J^{\pi^2}((RP, 1, 1, 1)) > 0 \;\; (3)$$

Juntando (2) e (3) temos que:

$$\forall \pi \in \Pi^S \;\; J^{\pi}(RP, 1, 1, 1) > 0 \Rightarrow J^{\pi^*}(RP, 1, 1, 1) > 0 \;\; (4)$$

Juntando (1) e (4) temos que:

$$\forall x \in \chi \;\; J^{\pi^*}(x) > 0$$

Concluímos que a frase *"For the MDP above, the cost-to-go function associated with the optimal policy is stricly positive"* é verdadeira.