

# MAC425/5739 - Inteligência Artificial

Vítor Kei Taira Tamada - 8516250

## Lista de exercícios 3

### Exercício 17.8 (Modificado) - Tomando decisões complexas

Questão a:

```
def valueIteration(mdp, e):
    S = mdp.s() # Conjunto de estados
    A = mdp.a() # Conjunto de ações
    R = mdp.r() # Conjunto de recompensas
    g = mdp.g() # Desconto gama do modelo de transição
    # mdp.p(fs, s, a) == P(s'|s, a)

    U = # Vetor de utilidades atual (U)
    Uf = # Vetor de utilidades futuro (U')
    delta = 0 # Mudança máxima da utilidade de qualquer estado em uma
iteração
    U_out = [] # Vetor que conterà o vetor de utilidades para cada valor
de r

    R11 = [-1000, -3, 0] # Valores que r (espaço R(1, 1) da grid world)
pode ter

    # Inicializa a utilidade de cada estado de S como zero
    for s in S:
        U[s] = 0
        Uf[s] = 0

    # Executa o algoritmo de iteração de valor com os 3 valores de R(1,
1) dados
    for rew in R11:
        R['(1, 1)'] = rew

        while delta < e * (1 - g) / g:
            U = Uf.copy() # U <- U'
            delta = 0 # delta <- 0

            for s in S:
                # U'[s] <- R(s) + gama * max{a in A(s)}(sum{s'}(P(s'|s, a)U[s']))
                sp = 0 # somatória de probabilidade
                for a in A[s]:
                    ue = 0 # utilidade esperada
                    for fs in S: # fs == future state == s'
                        ue += mdp.p(fs, s, a) * U[fs]
                    if ue > sp:
                        sp = ue
                sp *= g # gama * max(sum(P(s'|s, a)*U(s')))
                Uf[s] = R[s] + sp
```

```

        # if |U'[s] - U[s]| > delta then delta <- |U'[s] - U[s]|
        if abs(Uf[s] - U[s]) > delta:
            delta = abs(Uf[s] - U[s])

    U_out.append(U)

return U_out

```

■

Questão b:

**r = -1000**

→	→	+10
↓	→	↑
→	→	↑

$r = -1000$  implica na política acima pois estar na célula  $R(1,1)$  é tão ruim que só de ter uma possibilidade não-nula de cair nessa célula já torna a ação péssima. Portanto, as melhores ações são as de se afastar de  $R(1,1)$  o máximo possível e, ao mesmo tempo, tentar chegar em  $R(3,1)$ , dando preferência a se afastar de  $R(1,1)$ .

■

### Exercício 3 - P2 (2015)

Questão a:

$$V^i(S) = \max_A \sum_{S'} P(S'|S, A) [R(S, A, S') + \gamma V^{i-1}(S')]$$

$S$	$V^0$	$V^1(S)$
1	0	$0.80 \times (0 + 0.10 \times 0) + 0.20 \times (0 + 0.10 \times 0) = 0$
2	0	$0.80 \times (0 + 0.10 \times 0) + 0.20 \times (0 + 0.10 \times 0) = 0$
3	0	$0.80 \times (0 + 0.10 \times 0) + 0.20 \times (0 + 0.10 \times 0) = 0$
4	0	$0.80 \times (10 + 0.10 \times 0) + 0.20 \times (0 + 0.10 \times 0) = 8$
5	0	$0.80 \times (0 + 0.10 \times 0) + 0.20 \times (0 + 0.10 \times 0) = 0$
6	0	$0.60 \times (10 + 0.10 \times 0) + 0.40 \times (0 + 0.10 \times 0) = 6$

$S$	$V^1$	$V^2(S)$
1	0	$0.60 \times (0 + 0.10 \times 6) + 0.40 \times (0 + 0.10 \times 0) = 0.36$
2	0	$0.80 \times (0 + 0.10 \times 0) + 0.20 \times (0 + 0.10 \times 0) = 0$
3	0	$0.80 \times (0 + 0.10 \times 8) + 0.20 \times (0 + 0.10 \times 0) = 0.64$
4	8	$0.80 \times (10 + 0.10 \times 0) + 0.20 \times (0 + 0.10 \times 8) = 8.16$
5	0	$0.80 \times (0 + 0.10 \times 0) + 0.20 \times (0 + 0.10 \times 0) = 0$
6	6	$0.60 \times (10 + 0.10 \times 0) + 0.40 \times (0 + 0.10 \times 6) = 6.24$

$S$	$V^2$	$V^3(S)$
1	0.36	$0.60 \times (0 + 0.10 \times 6.24) + 0.40 \times (0 + 0.10 \times 0.36) = 0.39$
2	0	$0.80 \times (0 + 0.10 \times 0.64) + 0.20 \times (0 + 0.10 \times 0) = 0.06$
3	0.64	$0.80 \times (0 + 0.10 \times 8.16) + 0.20 \times (0 + 0.10 \times 0.64) = 0.67$
4	8.16	$0.80 \times (10 + 0.10 \times 0) + 0.20 \times (0 + 0.10 \times 8.16) = 8.17$
5	0	$0.80 \times (0 + 0.10 \times 0) + 0.20 \times (0 + 0.10 \times 0) = 0$
6	6.24	$0.60 \times (10 + 0.10 \times 0) + 0.40 \times (0 + 0.10 \times 6.24) = 6.25$

■

Questão b:

$$\pi^*(S) = \argmax_A \sum_{S'} P(S'|S, A) [R(S, A, S') + \gamma V^*(S')]$$

$$Q^*(A, S) = \sum_{S'} P(S'|S, A) [R(S, A, S') + \gamma V^*(S')]$$

$S$	$V^*(S)$	$Q^*(A = \text{horario}, S)$
1	0.39	$0.8 \times (0 + 0.1 \times 0.05) + 0.2 \times (0 + 0.1 \times 0.39) = 0.0118$
2	0.05	$0.8 \times (0 + 0.1 \times 0.67) + 0.2 \times (0 + 0.1 \times 0.05) = 0.0546$
3	0.67	$0.8 \times (0 + 0.1 \times 8.16) + 0.2 \times (0 + 0.1 \times 0.67) = 0.6662$
4	8.16	$0.8 \times (10 + 0.1 \times 0) + 0.2 \times (0 + 0.1 \times 8.16) = 8.1632$
5	0	$0.8 \times (0 + 0.1 \times 0) + 0.2 \times (0 + 0.1 \times 0) = 0$
6	6.25	$0.8 \times (0 + 0.1 \times 0.39) + 0.2 \times (0 + 0.1 \times 6.25) = 0.1562$

$S$	$V^*(S)$	$Q^*(A = \text{antihorario}, S)$
1	0.39	$0.6 \times (0 + 0.1 \times 6.25) + 0.4 \times (0 + 0.1 \times 0.39) = 0.3906$
2	0.05	$0.6 \times (0 + 0.1 \times 0.39) + 0.4 \times (0 + 0.1 \times 0.05) = 0.0254$
3	0.67	$0.6 \times (0 + 0.1 \times 0.05) + 0.4 \times (0 + 0.1 \times 0.67) = 0.0298$
4	8.16	$0.6 \times (0 + 0.1 \times 0.67) + 0.4 \times (0 + 0.1 \times 8.16) = 0.3666$
5	0	$0.6 \times (0 + 0.1 \times 0) + 0.4 \times (0 + 0.1 \times 0) = 0$
6	6.25	$0.6 \times (10 + 0.1 \times 0) + 0.4 \times (0 + 0.1 \times 6.25) = 6.2500$

$S$	$V^*(S)$	$Q^*(A = \textit{horario}, S)$	$Q^*(A = \textit{antihorario}, S)$	$\pi^*(S)$
1	0.39	0.0118	0.3906	$0.3906(AH)$
2	0.05	0.0546	0.0254	$0.0546(H)$
3	0.67	0.6662	0.0298	$0.6662(H)$
4	8.16	8.1632	0.3666	$8.1632(H)$
5	0	0	0	$0(H/AH)$
6	6.25	0.1562	6.2500	$6.2500(AH)$