

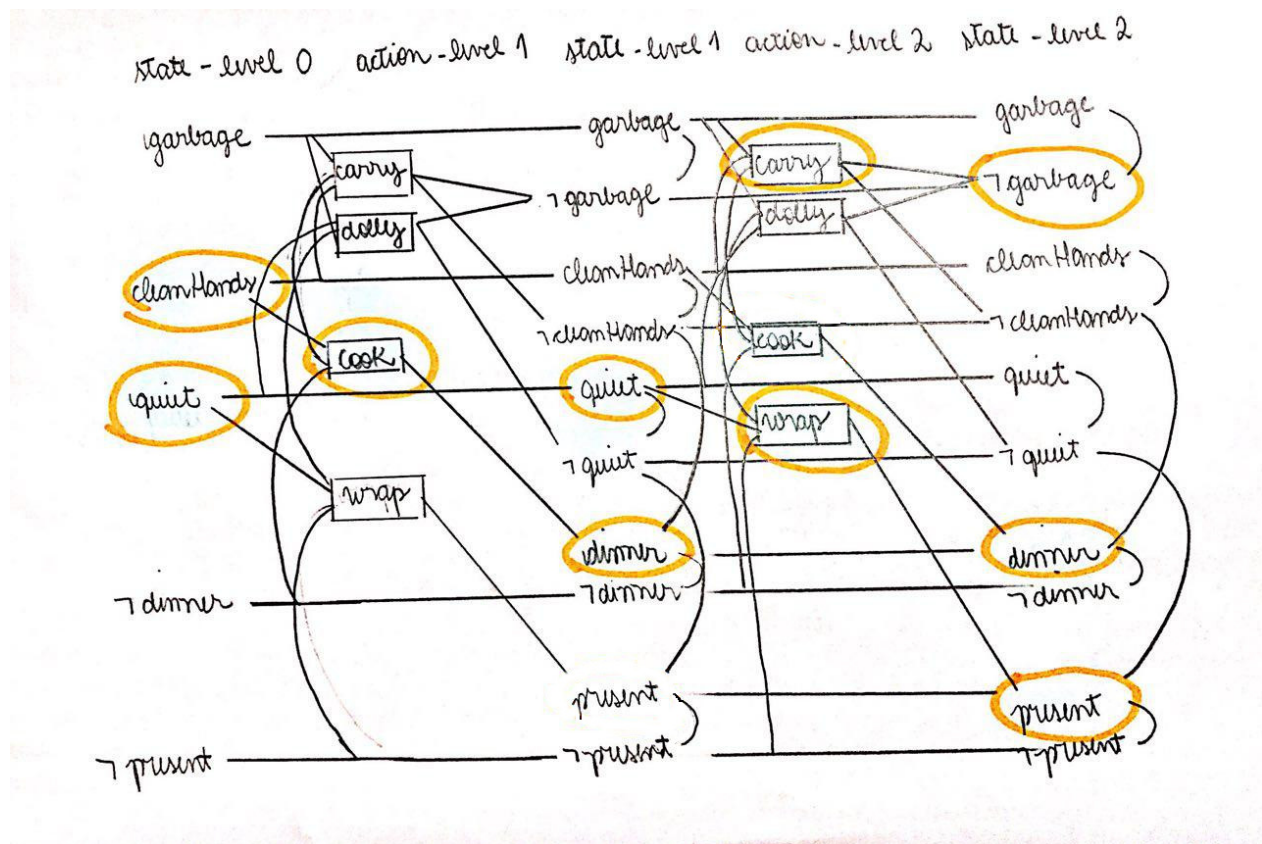
Nome: Luís Felipe de Melo Costa Silva
Número USP: 9297961

Lista de Exercícios 3 - MAC0425

Exercício 10.6

Quando os efeitos negativos de uma ação são descartados, nós adicionamos apenas os efeitos positivos, aumentando o conjunto de literais em que podemos encontrar a solução. Nesse caso, estamos simplificando as condições, tornando o problema menos restrito e mais geral, acabando com um problema mais fácil do que o original.

Exercício 10.9



Começamos o Graphplan com os seguintes literais e os que não temos da meta no *state-level 0*.

Então, aplicamos as ações possíveis a partir desses literais em *action-level 1*. Podemos observar os seguintes mutexes:

- Por interferência:
 - *carry* e *cook*
 - *dolly* e *wrap*
- Por efeitos inconsistentes:
 - *carry* e manutenção de *garbage*
 - *carry* e manutenção de *cleanHands*
 - *dolly* e manutenção de *garbage*
 - *dolly* e manutenção de *quiet*
 - *cook* e manutenção de $\neg dinner$
 - *wrap* e manutenção de $\neg present$

Chegamos ao *state-level 1*, e temos os seguintes mutexes:

- Por suporte inconsistente:
 - *garbage* e $\neg garbage$
 - *cleanHands* e $\neg cleanHands$
 - *quiet* e $\neg quiet$
 - *dinner* e $\neg dinner$
 - *present* e $\neg present$
- Por efeitos que competem
 - $\neg quiet$ e *present*
 - *garbage* e $\neg cleanHands$
 - *garbage* e $\neg quiet$
 - $\neg cleanHands$ e *dinner*

Prosseguimos para o *action-level 2*, e pode-se observar os mutexes:

- Por interferência e efeitos inconsistentes:

- manutenção de *garbage* e manutenção de \neg *garbage*
- manutenção de *cleanHands* e manutenção de \neg *cleanHands*
- manutenção de *quiet* e manutenção de \neg *quiet*
- manutenção de *dinner* e manutenção de \neg *dinner*
- manutenção de *present* e manutenção de \neg *present*
- Por interferência:
 - *carry* e *cook*
 - *dolly* e *wrap*
- Por efeitos inconsistentes:
 - *carry* e manutenção de *garbage*
 - *carry* e manutenção de *cleanHands*
 - *dolly* e manutenção de *garbage*
 - *dolly* e manutenção de *quiet*
 - *cook* e manutenção de \neg *dinner*
 - *wrap* e manutenção de \neg *present*
- Por necessidades que competem:
 - *cook* e manutenção de \neg *cleanHands*
 - *wrap* e manutenção de \neg *quiet*

Por fim, temos o *state-level 2*, que nos dá os seguintes mutexes:

- Por suporte inconsistente:
 - *garbage* e \neg *garbage*
 - *cleanHands* e \neg *cleanHands*
 - *quiet* e \neg *quiet*
 - *dinner* e \neg *dinner*
 - *present* e \neg *present*
- Por efeitos que competem

- $\neg quiet$ e $present$
- $garbage$ e $\neg cleanHands$
- $garbage$ e $\neg quiet$
- $\neg cleanHands$ e $dinner$

Para encontrarmos o plano para resolver esse problema, começamos selecionando os literais da meta em *state-level 2*. Então, pegamos ações no *action-level 2* que não estão em mutex entre si (*carry*, *wrap*) e as pré-condições dessas ações (*quiet*) e a ação de manutenção dos literais que não foram contemplados com as ações escolhidas (*dinner*), que estão no *state-level 1*. Olhamos para o *action-level 1* e procuramos as ações que geram *dinner* e *quiet*. A ação *cook* gera *dinner*, e *quiet* é gerado por uma ação de manutenção. Logo, nosso plano é $[cook, carry, wrap]$. Existem outros planos possíveis, por exemplo: $[wrap, cook, dolly]$.

Exercício MDP

Usando a definição de que MDP é uma tupla $\langle S, D, A, T, R \rangle$ do slide 3 da segunda aula de Planejamento Probabilístico, usaremos a seguinte fórmula para calcular V^n (do slide 14 da mesma aula):

$$V^n(s) = R(s) + \max_{a \in A} (\sum_{s' \in S} T(s, a, s') \cdot V^{n-1}(s'))$$

Começamos com $V^0(s) = R(s), \forall s \in S$. Então, teremos:

- Para $n = 1$

- $V^1(s1) = 0 + \max(V^0(s2); 0,5 \cdot V^0(s1) + 0,5 \cdot V^0(s4))$
 $= \max(0; 0,5 \cdot 0 + 0,5 \cdot 100) = 50$
- $V^1(s2) = 0 + \max(0,8 \cdot V^0(s3) + 0,2 \cdot V^0(s5))$
 $= \max(0,8 \cdot 0 + 0,2 \cdot (-100)) = -20$
- $V^1(s3) = 0 + \max(V^0(s2); V^0(s4))$
 $= \max(0; 100) = 100$
- $V^1(s4) = 100 + \max(V^0(s1); V^0(s3); V^0(s5))$
 $= 100 + \max(0; 0; -100) = 100$

$$\begin{aligned}
- V^1(s5) &= -100 + \max(V^0(s2); V^0(s4)) \\
&= -100 + \max(0; 100) = 0
\end{aligned}$$

• Para $n = 2$

$$\begin{aligned}
- V^2(s1) &= 0 + \max(V^1(s2); 0,5 \cdot V^1(s1) + 0,5 \cdot V^1(s4)) \\
&= \max(-20; 0,5 \cdot 50 + 0,5 \cdot 100) = 75 \\
- V^2(s2) &= 0 + \max(0,8 \cdot V^1(s3) + 0,2 \cdot V^1(s5)) \\
&= \max(0,8 \cdot 100 + 0,2 \cdot 0) = 80 \\
- V^2(s3) &= 0 + \max(V^1(s2); V^1(s4)) \\
&= \max(-20; 100) = 100 \\
- V^2(s4) &= 100 + \max(V^1(s1); V^1(s3); V^1(s5)) \\
&= 100 + \max(50; 100; 0) = 200 \\
- V^2(s5) &= -100 + \max(V^1(s2); V^1(s4)) \\
&= -100 + \max(-20; 100) = 0
\end{aligned}$$

• Para $n = 3$

$$\begin{aligned}
- V^3(s1) &= 0 + \max(V^2(s2); 0,5 \cdot V^2(s1) + 0,5 \cdot V^2(s4)) \\
&= \max(80; 0,5 \cdot 75 + 0,5 \cdot 200) = 137,5 \\
- V^3(s2) &= 0 + \max(0,8 \cdot V^2(s3) + 0,2 \cdot V^2(s5)) \\
&= \max(0,8 \cdot 100 + 0,2 \cdot 0) = 80 \\
- V^3(s3) &= 0 + \max(V^2(s2); V^2(s4)) \\
&= \max(80; 200) = 200 \\
- V^3(s4) &= 100 + \max(V^2(s1); V^2(s3); V^2(s5)) \\
&= 100 + \max(75; 100; 0) = 200 \\
- V^3(s5) &= -100 + \max(V^2(s2); V^2(s4)) \\
&= -100 + \max(80; 200) = 100
\end{aligned}$$

As tabelas V e π ficam assim:

s1	0	50	75	137,5	s1	move(r1, l1, l4)	move(r1, l1, l4)	move(r1, l1, l4)
s2	0	-20	80	80	s2	move(r1, l2, l3)	move(r1, l2, l3)	move(r1, l2, l3)
s3	0	100	100	200	s3	move(r1, l3, l4)	move(r1, l3, l4)	move(r1, l3, l4)
s4	100	100	200	200	s4	move(r1, l4, l3)	move(r1, l4, l3)	move(r1, l4, l3)
s5	-100	0	0	100	s5	move(r1, l5, l4)	move(r1, l5, l4)	move(r1, l5, l4)

Exercício Q-Learning

Para esse exercício vamos usar a fórmula do *Q-value update*, que consiste em:

$$Q(s, a) \leftarrow (1 - \alpha) \cdot Q(s, a) + \alpha(r + \gamma \cdot \max_{a' \in A} Q(s', a'))$$

O enunciado nos pede para usar $\alpha = 1$ e $\gamma = 0,9$. Com esses valores, nossa fórmula ficará:

$$Q(s, a) \leftarrow (r + 0,9 \cdot \max_{a' \in A} Q(s', a'))$$

Começando com a tabela zerada e fazendo os cálculos para os episódios pedidos, teremos:

1.

$$Q(3, S) = -10 + 0,9 \cdot 0 = -10$$

$$Q(3, N) = 0 + 0,9 \cdot 0 = 0$$

$$Q(1, N) = -10 + 0,9 \cdot 0 = -10$$

$$Q(1, O) = 10 + 0,9 \cdot 0 = 10$$

	N	S	L	O
0	0	0	0	0
1	-10	0	0	10
2	0	0	0	0
3	0	-10	0	0

2.

$$Q(2, L) = 0 + 0,9 \cdot 0 = 0$$

$$Q(3, N) = 0 + 0,9 \cdot 0 = 0$$

$$Q(1, N) = -10 + 0,9 \cdot 0 = -10$$

$$Q(1, L) = -10 + 0,9 \cdot 0 = -10$$

$$Q(1, O) = 10 + 0,9 \cdot 0 = 10$$

	N	S	L	O
0	0	0	0	0
1	-10	0	-10	10
2	0	0	0	0
3	0	0	0	0