

Aprendizagem e Decisão Inteligente		2017/2018
Filipe Rafael Soares	76543	Grupo 59
Ana Luísa Santo	79758	HW-3

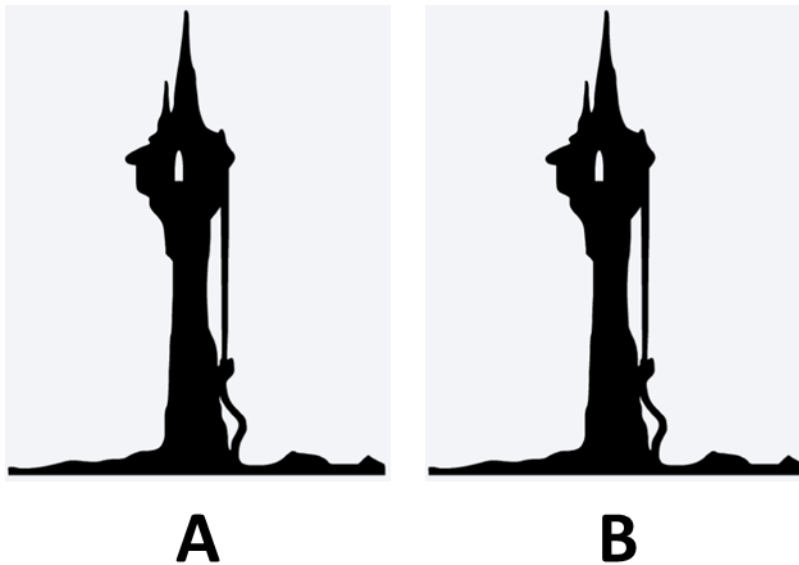


Figura 1 – Torre A e Torre B são hipóteses que o cavaleiro tem de escolher entre (cenário em discussão neste documento).

1)

Neste trabalho irá ser modelado a decisão do cavaleiro entre escolher a Torre A (T_a) ou a Torre B (T_b) de modo a resgatar a princesa, sobre forma de um POMDP. O cavaleiro pode fazer a ação de invadir a torre A (inv_a) ou de invadir a torre B (inv_b), sendo que também tem a hipótese de espreitar (**peek**). O espaço de observações acaba por ser igual ao espaço de estado, uma vez que vemos se a princesa está na torre A ou na torre B, acrescentando o facto de poder não ter havido observações (\emptyset)

Assim sendo obtemos o seguinte:

$$\mathcal{X} = \{T_a \quad T_b\}$$

$$\mathbf{A} = \{inv_a, inv_b, peek\}$$

$$\mathcal{Z} = \{T_a \quad T_b \quad \emptyset\}$$

2)

Para cada ação existe uma matriz de transição associada a esta, havendo assim as matrizes de transição de invadir a Torre A (P_{inva}), invadir a Torre B (P_{invb}) e espreitar (P_{peek}). Estas têm os seguintes valores:

$$P_{inva} = P_{invb} = \begin{bmatrix} 0,5 & 0,5 \\ 0,5 & 0,5 \end{bmatrix} \quad P_{peek} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Dado que quando o cavaleiro escolhe uma torre, o jogo recomeça de forma aleatória e a torre a escolher na próxima jogada poderá ser qualquer uma delas com igual probabilidade, isto faz com que P_{inva} e P_{invb} tenham os mesmos valores. Em relação a P_{peek} , uma vez que quando o cavaleiro espreita, está ação não influencia o instante em que foi realizada, apenas o instante imediatamente a seguir, podemos ter certeza que escolhendo a ação peek, $\mathcal{X}_{t+1} = \mathcal{X}_t$ com probabilidade 1.

No que diz respeito às matrizes de probabilidade observação, estas são as seguintes:

$$O_{inva} = O_{invb} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad O_{peek} = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0.1 & 0.9 & 0 \end{bmatrix}$$

A matriz associada à observação da ação de espreitar (O_{peek}) é referida no enunciado pelo que não necessita de qualquer explicação. A matriz associada à observação da ação de invadir a torre A ou a torre B (O_{inva}/O_{invb}), tendo em conta o que já foi mencionado anteriormente e acrescentando o facto de que da ação apenas podem acontecer 2 cenários (ou o cavaleiro salva a princesa ou é expulso da torre), faz com que a matriz seja preenchida com os valores apresentados.

Por último a matriz de custos pode ser a seguinte:

$$c = \begin{bmatrix} 0 & 1 & 0.3 \\ 1 & 0 & 0.3 \end{bmatrix}$$

Passando a explicar, pretende-se que quando o cavaleiro realize a ação de invadir a torre A, estando a princesa na torre A, não tenha qualquer custo, caso contrário tenha a penalização máxima e por analogia para torre B. A ação de espreitar tem um custo mínimo para que este nem sempre opte por realizar ação e assim se aproxime mais do real.

3)

Usando a regra de atualização da belief, optemos:

$$b_{new} = \mu b_{old} P_{peek} \text{diag}(O_{peek}) = [0.7 \quad 0.3] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix} = [0.206 \quad 0.794]$$

Onde μ é um fator que normaliza a nova belief.