

Instituto Superior de Engenharia de Lisboa

Departamento de Engenharia de Electrónica e Telecomunicações e de Computadores
Licenciatura em Engenharia de Redes de Comunicação e Multimédia

Inteligência Artificial para Sistemas Autónomos 2011/12

EXAME DE 1ª ÉPOCA

**Responda às questões apresentadas de forma clara e objectiva,
justificando sempre que necessário.**

1. Considere as arquitecturas de agente anteriormente estudadas.
 - 1.1. Qual o papel da coordenação comportamental no contexto de uma arquitectura de agentes reactivos? Justifique.
 - 1.2. Qual o papel da memória numa arquitectura de agentes reactivos? Justifique.
 - 1.3. Como é concretizada a noção de racionalidade numa arquitectura de agente? Justifique.
2. Pretende-se implementar um agente capaz de operar num ambiente caracterizado por um conjunto de estados S , por um conjunto de operadores O e pela função de transição de estado T , de seguida indicados.

$S = \{s_0, s_1, s_2, s_3, s_4, s_5\}$

$O = \{a_0, a_1, a_2, a_3\}$

$T(s_0, a_0) = s_1$

$T(s_1, a_1) = s_4$

$T(s_4, a_0) = s_5$

$T(s_4, a_2) = s_3$

$T(s, a) = s$, nas restantes situações

Inicialmente o agente encontra-se no estado **s0** e tem por objectivo atingir o estado **s4**. Por omissão, têm maior prioridade os estados com maior índice.

- 2.1. Tendo por base o método de *procura em profundidade limitada* (limite de profundidade = 3), elabore e apresente a árvore de procura, bem como as respectivas estruturas de dados auxiliares. Indique a solução obtida.
- 2.2. Tendo por base o método de *procura em largura*, elabore e apresente a árvore de procura, bem como as respectivas estruturas de dados auxiliares. Indique a solução obtida.
- 2.3. Relacione os métodos de *procura em profundidade* e *procura em profundidade iterativa*, indicando as principais características de cada método, bem como as condições que devem ser garantidas para que essas características se verifiquem.

3. Pretende-se implementar um agente capaz de operar num ambiente descrito pelo modelo de seguida apresentado.

$S = \{ s_0, s_1, s_2, s_3, s_4, s_5 \}$	$T(s_0, a_0, s_1) = 0.8$
$A = \{ a_0, a_1, a_2, a_3 \}$	$T(s_0, a_0, s_3) = 0.2$
$R(s, a, s') = -1$, se $s' = s_2$	$T(s_1, a_1, s_4) = 0.8$
$R(s, a, s') = -1$, se $s' = s_3$	$T(s_1, a_1, s_2) = 0.2$
$R(s, a, s') = +1$, se $s' = s_5$	$T(s_4, a_0, s_5) = 0.8$
	$T(s_4, a_0, s_3) = 0.2$
	$T(s_4, a_2, s_3) = 1$
$R(s, a, s') = -0.1$, nas restantes situações	$T(s, a, s') = 0$, nas restantes situações

Considere uma abordagem baseada em processos de decisão de Markov, e um factor de desconto temporal $\gamma = 1$.

- 3.1. Tendo por base uma política de selecção de acção *greedy*, resolva o processo de decisão de Markov correspondente à representação anterior, com um limiar de convergência de 0.5 ou até ao máximo de 3 iterações.
 - 3.2. Represente a política comportamental resultante da resolução da alínea anterior.
 - 3.3. Qual o papel da função de utilidade num processo de decisão de Markov? Justifique.
4. Pretende-se implementar um agente capaz de aprender, por reforço, a agir no ambiente descrito na questão anterior. Considere um agente no estado inicial $s = s_0$, $\alpha = 0.5$, $\gamma = 0.9$, $\varepsilon = 0$. Por omissão o agente realiza a acção com menor índice das acções possíveis.
- 4.1. Tendo por base o método de aprendizagem *Q-Learning* e uma política de selecção de acção ε -*greedy*, simule a execução do algoritmo indicando para cada iteração os estados, acções e recompensas envolvidas, bem como os valores da função $Q(s,a)$ alterados em cada iteração, para 3 iterações de aprendizagem.
 - 4.2. Represente a política comportamental resultante da resolução da alínea anterior.
 - 4.3. Relacione os métodos de aprendizagem *Q-Learning* e *Dyna-Q*. Existe alguma vantagem relativa entre estes métodos? Justifique.