CS781 Quiz 3 (Autumn 2023)

Max marks: 35 **Duration: 60 mins**

- *The exam is open book and notes. However, you are not allowed to search on the internet or consult others over the internet for your answers.*
- *Be brief, complete and stick to what has been asked.*
- *Unless asked for explicitly, you may cite results/proofs covered in class without reproducing them.*
- ***If you need to make any assumptions, state them clearly.***
- ***Do not copy solutions from others. Penalty for offenders: FR grade.***

1. Consider an agent interacting with an environment modeled by the MDP $M = (S, A, P, \eta)$, where the set of states $S = \{q_0, q_1, q_2\}$, the set of actions $A = \{X, Y\}$, the transition probabilities P are as shown in the table below, and the initial state probability distribution is $\eta(q_0) = \eta(q_2) = 0.5$.

s_i	a_i	s_{i+1}	$P(s_i, a_i, s_{i+1})$	s_i	a_i	s_{i+1}	$P(s_i, a_i, s_{i+1})$	s_i	a_i	s_{i+1}	$P(s_i, a_i, s_{i+1})$
q_0	X	q_0	0.5	q_0	X	q_1	0.5	q_0	X	q_2	0
q_0	Y	q_0	0.5	q_0	Y	q_1	0	q_0	Y	q_2	0.5
q_1	X	q_0	0.5	q_1	X	q_1	0	q_1	X	q_2	0.5
q_1	Y	q_0	0.5	q_1	Y	q_1	0	q_1	Y	q_2	0.5
q_2	X	q_0	0	q_2	X	q_1	0.5	q_2	X	q_2	0.5
q_2	Y	q_0	0.5	q_2	Y	q_1	0.5	q_2	Y	q_2	0

Assume the agent can see the states of the MDP as it interacts with the environment. However, because of limited learning capabilities, the agent can learn only one of two deterministic posititional policies π_1 or π_2, described by the table below. Recall that a deterministic positional policy maps a finite trajectory $s_0\, a_0\, s_1 \cdots s_{k-1}\, a_{k-1}\, s_k$ of MDP states s_i and actions a_i to a unique next action a_k depending only on s_k.

s_k	a_k (policy π_1)	a_k (policy π_2)
q_0	Y	X
q_1	X	Y
q_2	X	Y

The agent is required to learn a (compositional) policy for the task described by the following specification in SPECTRL:
`achieve(reach p1) ; (achieve(reach p0) or achieve(reach p2)); achieve(reach p3)`,
where predicates $p0, p1, p2$ and $p3$ correspond to sets of MDP states $\{q_0\}$, $\{q_1\}$, $\{q_2\}$ and $\{q_1, q_2\}$ respectively.

   (a) *[5 marks]* Draw the abstract graph $\mathcal{G} = (U, E, u_0, F, \beta, \mathcal{Z}_{safe})$ for the above specification, clearly indicating all components. For the sets of traces corresponding to various edges in E, you can use a textual description to describe these sets, if needed.

   (b) *[10 marks]* If we apply the DIRL algorithm to compute the compositional policy of the agent, which of π_1 or π_2 would be chosen as the edge policy for each edge going out of the initial vertex u_0 in the above graph? Clearly justify your answer.

(c) *[10 marks]* What would the edge weight of each edge going out of u_0 be, if we use the edge policies determined above?

(d) *[10 marks]* Suppose we use policy π_1 for all edges in the abstract graph. Give as good an estimate as you can of the probability of the agent satisfying the specification given in SPECTRL.