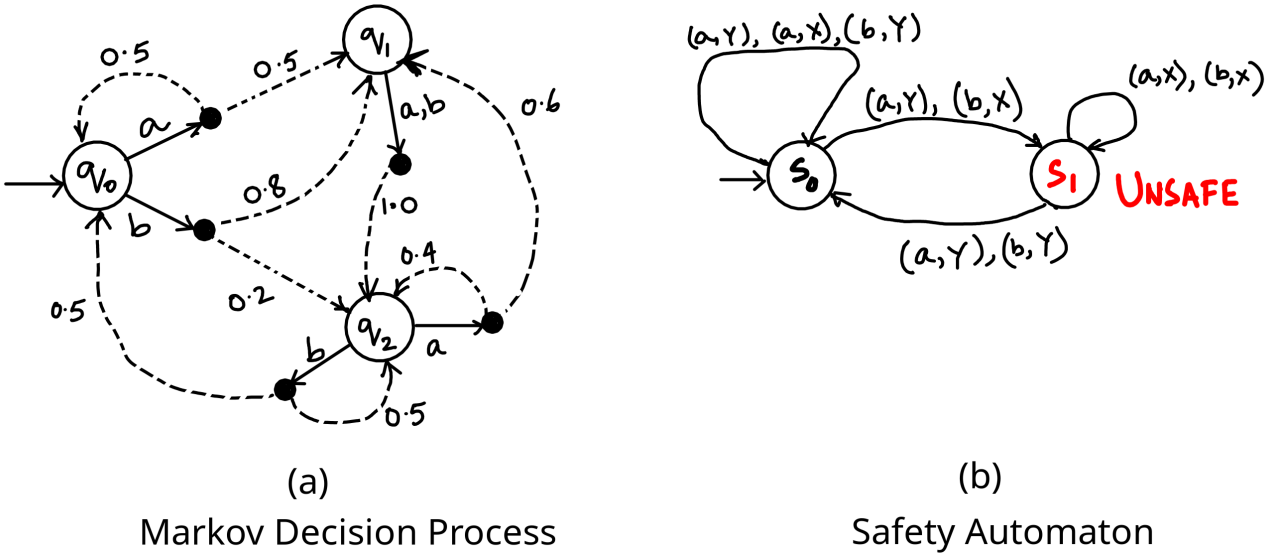
CS781 Quiz 2 (Autumn 2024)

Max marks: 25 **Duration: 90 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 the MDP shown in Fig. 1(a) representing interactions between an RL agent and the environment. The set of agent actions in is $\{a, b\}$. The probabilistic moves of the environment are represented by dashed arrows in the MDP. Assume the observation function for MDP states is given by $f : \{q_0, q_1, q_2\} \to \{X, Y\}$, where $f(q_0) = f(q_2) = X$ and $f(q_1) = Y$.



(a) Markov Decision Process

(b) Safety Automaton

Figure 1: MDP and safety automaton

A *non-deterministic* safety automaton for the agent-environment is shown in Fig. 1(b). When the MDP is in state q_i and the agent takes action $\alpha \in \{a, b\}$, the safety automaton makes a non-deterministic transition on the input $(\alpha, f(q_i))$. For example, if the sequence of states and actions of the MDP is $q_0\ a\ q_1\ a\ q_2\ a\ q_2\ a \ldots$, the safety automaton moves on the sequence of inputs $(a, f(q_0)), (a, f(q_1), (a, f(q_2), (a, f(q_2)), \ldots$, i.e. on the input sequence $(a, X), (a, Y), (a, X), (a, X), \ldots$. Note that the automaton is non-deterministic, so it can have multiple runs on the above input sequence. Here are two possible runs:

- Run 1: $S_0 \xrightarrow{(a,X)} S_0 \xrightarrow{(a,Y)} S_0 \xrightarrow{(a,X)} S_0 \xrightarrow{(a,X)} S_0 \xrightarrow{\cdots} \ldots$
- Run 2: $S_0 \xrightarrow{(a,X)} S_0 \xrightarrow{(a,Y)} S_1 \xrightarrow{(a,X)} S_1 \xrightarrow{(a,X)} S_1 \xrightarrow{\cdots} \ldots$

State S_1 is the unsafe state in the automaton. As long as the automaton has even a single run that encounters the unsafe state, we consider the interaction of the agent and environment to be unsafe. Thus, the above sequence of states and actions of the MDP will be deemed unsafe (run 2 of the safety automaton passed through the unsafe state S_1).

(a) *[3 marks]* Let A_M be a *non-deterministic* automaton on the alphabet $\{a, b\} \times \{X, Y\}$ that *abstracts* the behaviour of the MDP. You are allowed to have upto 3 states of A_M, *including the state that is reached on input sequences that represent infeasible runs of the MDP*. Note that this means you cannot simply use q_0, q_1, q_2 from the MDP as the feasible states of A_M. You can, however, effectively "merge" two states of the MDP to obtain one state of A_M, while making sure that the incoming and outgoing transitions of this "merged" state are appropriately constructed. Give the automaton A_M obtained by effectively merging states q_0 and q_2 of the MDP.

(b) *[3 marks]* Using the automaton A_M obtained above and the safety automaton given in Fig. 1(b), construct the game graph (cross-product automaton), clearly marking all unsafe nodes in the graph.

(c) *[4 marks]* Identify the winning region (set of game-graph nodes) in the above game graph. You must show all steps and reasoning clearly, else you will not get any marks. Simply stating the winning region will fetch no marks.

(d) *[5 marks]* Is it possible to construct a pre-emptive non-probabilistic shield in this case?
If your answer is in the negative, state reasons why it is impossible to construct such a shield. If your answer is in the positive, show the steps of construction of the pre-emptive shield, and give the final shield. Note that a pre-emptive shield allows the agent all possible safe actions at every step, so that the agent can choose the best one (according to its reward function) among these.

2. A super-secret decision tree (DT) is used by a bank to predict the eligibility of loans based on an applicant's age (a), salary (s) and credit rating (c). A bank employee has access to the DT as a *blackbox*, i.e. she can feed in triples (a, s, c) and obtain a "Y"/"N" answer. Being impressed by the predictions of the blackbox DT, she sets out one afternoon to find abductive explanations of some loan eligibility predictions. However, she has no idea what the actual DT looks like or what predicates are used in it. So, she *guesses* that the predicates used are possibly a_{50} (age $\geq$ 50?), s_{150} (salary $\geq$ Rs. 150K) and c_{80} (creding rating $\geq$ 80). Since she does not have access to the formula φ representing the DT (recall this is needed for the abductive explanation generation algorithm studied in class), she decides to use the following dataset that she processed that afternoon as satisfying assignments of φ: $\{(50, 120, 80, Y), (45, 80, 60, N), (35, 160, 40, N), (60, 170, 90, Y), (55, 200, 60, Y)\}$, where each tuple represents (age, salary, credit-ranting, decision).

(a) *[2 × 3 marks]* For each of the following datapoints, indicate a cardinality-minimal abductive explanation in terms of a_{50}, s_{150}, c_{80} that the employee can get.

i. $(50, 120, 80, Y)$

ii. $(55, 200, 60, Y)$

(b) *[4 marks]* From the answer to the previous sub-question, can you construct a decision tree (DT) using the predicates a_{50}, s_{150}, c_{80} that serves to explain all the five datapoints. If it is not possible to construct such a DT, explain why. Otherwise, give such a DT.