BITS Pilani, K. K. Birla Goa Campus

Artificial Intelligence (CS F407)

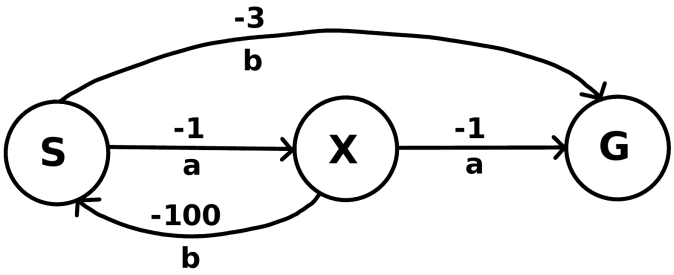
Comprehensive Exam (19/12/2022)
Total Marks: 40
Time Limit: 3 Hours

This question paper contains **5** questions carrying **40** marks. All questions are compulsory. Answer each question on a fresh page. Answer all the parts of a question in the same place. Show the necessary steps and justifications for your answers.

Control Algorithms in Reinforcement learning

Question 1 (10 marks)

Consider the Markov Decision Process shown below:



State S is the starting state and state G is the terminal state. Each state has two possible actions : a and b. Action values for the terminal state (i.e. $Q(G, \cdot)$) are initialized to 0. The remaining action values are initialized to arbitrary values. The discount rate (γ) is .8.

(a) (3 marks) Suppose we use Sarsa on-policy TD control algorithm to estimate the action values. If the algorithm keeps simulating episodes, what will the following two estimates converge to:

(i) Q(S,a) (ii) Q(X,b)

Round your answers to **four** decimal places. Assume that the algorithm uses an ϵ-greedy behavior policy, where $\epsilon = .1$ (a constant). Also, assume that the step rate α is very small so that the algorithm converges.

(b) (3 marks) Suppose we use Q-learning off-policy TD control algorithm to estimate the action values. If the algorithm keeps simulating episodes, what will the following two action value estimates converge to:

(i) Q(S,a) (ii) Q(X,b)

Round your answers to **four** decimal places. Assume that the algorithm uses an ϵ-greedy behavior policy, where $\epsilon = .1$. Also, assume that the step rate α is small.

(c) (2 marks) Online performance (average sum of rewards during an episode) for which of the two algorithms mentioned above will be better? Why?

(d) (2 marks) Which of the two algorithms mentioned above will converge to the optimal action values faster? Why?

Reasoning Under Uncertainty

Question 2 (8 marks)

On a sunny day ($Sunny$), a person may either go for a $Picnic$ or for a $Movie$ (or both) with certain probabilities. $Sunny$, $Picnic$ and $Movie$ are random variables that take either $True$ or $False$ values. Assume that the two random variables $Movie$ and $Picnic$ are conditionally independent given variable $Sunny$. The full joint distribution table for $Sunny$, $Picnic$ and $Movie$ is as shown below.

	$picnic$		$\neg picnic$	
	$movie$	$\neg movie$	$movie$	$\neg movie$
$sunny$	.252	.168	.108	.072
$\neg sunny$	.005	x	.095	y

(a) (3 marks) Find the values x and y. Show the necessary steps and justifications. Round your answer to three decimal places.

(b) (2 marks) Suppose it is given that the person did *not* go for a movie. What is the probability that it was sunny that day? Round your answer to four decimal places.

(c) (3 marks) Use the conditional independence property mentioned above to construct a Bayesian network that can represent the above joint probability distribution table in a *concise* manner. Show the conditional probabilities associated with each node in the Bayesian network.

Knowledge representation using First-order logic

Question 3 (7 marks)

Suppose we want to construct a knowledge base (KB) using first-order logic (FOL). We have unary predicate $State$, binary predicate $Borders$ and binary predicate In. For example, $State(S)$ represents S is a state; $Borders(S, Karnataka)$ represents S borders Karnataka; and $In(S, SouthIndia)$ represents S is in South India.

Suppose we want to represent the following facts in first order logic using the predicates described above:

R1. *Every state in South India borders Karnataka.*

R2. *Punjab and Telangana are states.*

R3. *Telangana borders Karnataka.*

R4. *Punjab does not border Karnataka.*

(a) (3 marks) Construct a knowledge base (KB) that represents the above facts using FOL sentences. Write the sentences in conjunctive normal form.

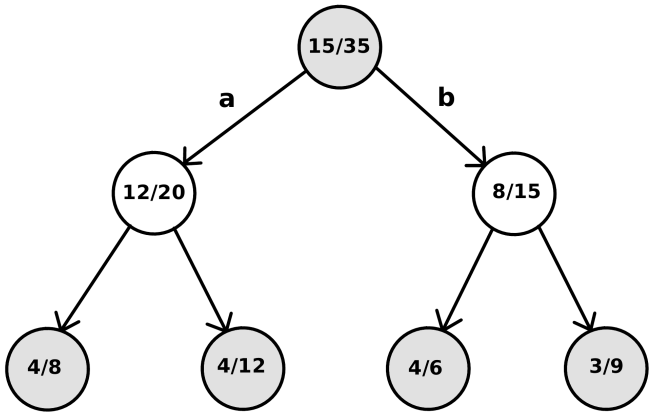
(b) (4 marks) How will the resolution algorithm check whether the following entailments hold? List all the *new* resolvent clauses that the resolution algorithm will derive while checking each of the following two entailments. Mention the unifier where necessary.

(i) $KB \models In(Telangana, SouthIndia)$ (ii) $KB \models \neg In(Punjab, SouthIndia)$

Monte Carlo Tree Search

Question 4 (8 marks)

We have a two-player, turn-taking, zero-sum game where the best action is chosen using the Monte Carlo Tree Search (MCTS) algorithm. The figure below shows the search tree for the MCTS:



Assume that the game has two players : white and black. In each game state, the next player can choose between two possible actions. The root node (level-1) represents a state where white player has just moved and white has won 15 out of 35 playouts. The level-2 nodes show the number of wins and playouts for the black player (e.g. black has won 12 out of 20 playouts at the left level-2 node). At the root node, the black player needs to make the next move using the MCTS algorithm. Consider the following two iterations of the MCTS algorithm for selecting the best action at the root node:

Iteration 1 : Selection, Expansion, Simulation (Black player *loses*), Back-propagation.

Iteration 2 : Selection, Expansion, Simulation (Black player *wins*), Back-propagation.

Assume that exactly one new node is generated in each Expansion step. The selection policy is based on the upper confidence bounds applied to trees (UCB). The UCB value of node n is as given below:

$$UCB(n) = \frac{U(n)}{N(n)} + \sqrt{\frac{2 \times \log_2 N(Parent(n))}{N(n)}}$$

(a) (2 marks) Draw the game tree that we will get after the first iteration of the MCTS algorithm. Mark the wins and playouts within each node of the game tree.

(b) (2 marks) Draw the game tree that we will get after the second iteration of the MCTS algorithm. Mark the wins and playouts within each node of the game tree.

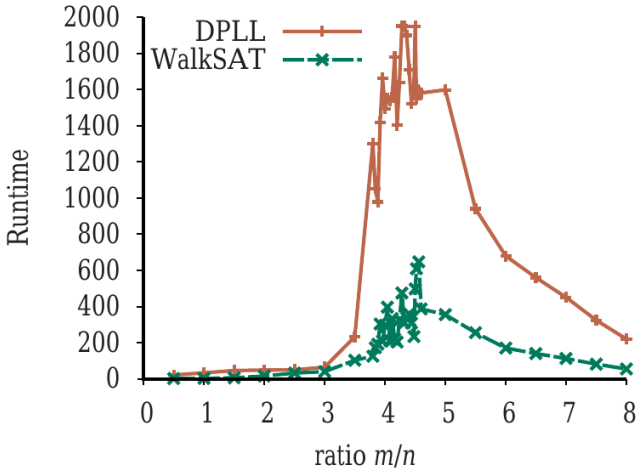
(c) (2 marks) Give two reasons why heuristic alpha-beta tree search is not effective for a game like Go?

(d) (2 marks) How does Monte Carlo Tree search overcome the difficulties faced by the heuristic alpha-beta tree search for the game Go?

Propositional logic inference using SAT solvers

Question 5 (7 marks)

We have seen the graph shown below while discussing the DPLL and WalkSAT algorithms.



(a) (1 mark) In the above graph, what does ratio m/n mean? What is m? What is n?

(b) (2 marks) In the above graph, why is the running time of WalkSAT algorithm low when the ratio m/n is around 1 and high when the ratio is around 4.5?

(c) (2 marks) Why is the running time of DPLL algorithm high when the ratio m/n is around 4.5 and low when the ratio is around 8?

(d) (2 marks) Consider the propositional logic sentence given below:

$$c \wedge (b \vee e \vee \neg d) \wedge (a \vee \neg f \vee e) \wedge (d \vee \neg c \vee \neg a) \wedge (c \vee \neg e \vee \neg d) \\ \wedge (f \vee \neg b) \wedge (d \vee \neg f \vee a) \wedge (\neg c \vee \neg d) \wedge (d \vee \neg a \vee \neg b)$$

Suppose we have an algorithm $\mathcal{A}$ that repeatedly applies two (DPLL algorithm) heuristics — Unit clause heuristic and Pure symbol heuristic. The algorithm $\mathcal{A}$ alternates between applying the Unit clause and Pure symbol heuristic in each iteration. In each iteration, only one of the two heuristics is used; And exactly one symbol is assigned a truth value. In the first iteration, the algorithm starts with applying the Unit clause heuristic. Algorithm $\mathcal{A}$ terminates when either all the symbols are assigned values or when the heuristic for the i^{th} iteration cannot be applied. Find the sequence in which algorithm $\mathcal{A}$ will assign truth values to the symbols for the sentence given above. You should also mention the truth value assigned for each symbol.

(Note: Your answer should include only those symbols that are assigned a truth value before algorithm $\mathcal{A}$ terminates. There will be no partial marking for this question.)