THE UNIVERSITY OF HONG KONG

FACULTY OF ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE

COMP7404 Computational Intelligence and Machine Learning

Date: May 20, 2014 **Time: 6:30pm - 8:30pm**

Only approved calculators as announced by the Examinations Secretary can be used in this examination. It is the candidates' responsibility to ensure that their calculator operates satisfactorily, and candidates must record the name and type of the calculator used on the front page of the examination script.

Calculator Brand: ____________________ *Calculator Model:* ____________________

Answer all six questions in the space provided.

Special Note: Candidates are permitted to bring to the examination one sheet of A4-sized paper with hand-written notes on both sides.

University No.: ________________________________

Question No.	Weight (in %)	Your Mark (in %)
1	10	
2	25	
3	10	
4	20	
5	25	
6	10	
Total	100	

1. Introduction

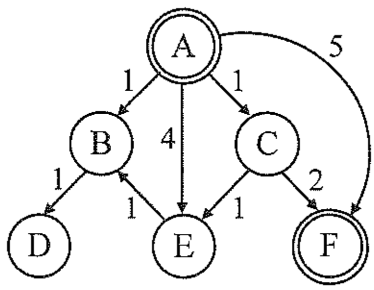
(a) Write down a brief and precise definition of the following. *(10 marks)*

i. Performance measure

ii. Stochastic environment

2. Search

(a) Consider the following search graph in which A is the start state and F is the goal state. Assume that for each search method, the details are such that a state's children will be visited in alphabetical order when possible. *(7 marks)*



i. Write down the order in which states will be expanded by depth-first tree search.

ii. Write down the order in which states will be expanded by a breadth-first graph search.

iii. Write down the order in which states will be expanded by uniform-cost graph search.

(b) Consider the problem of moving k knights from k starting squares $s_1, \ldots, s_k$ to k goal squares $g_1, \ldots, g_k$, on an unbounded chessboard, subject to the rule that no two knights can land on the same square at the same time. Each action consists of moving up to k knights simultaneously. We would like to complete the maneuver in the smallest number of actions. *(10 marks)*

i. What is the maximum branching factor in this state space, expressed as a function of k?

ii. Suppose h_i is an admissible heuristic for the problem of moving knight i to goal g_i by itself. Which of the following heuristics are admissible for the k-knight problem? Of those, which is the best? Explain.

A. $min\{h_1, h_2, \ldots, h_k\}$

B. $max\{h_1, h_2, \ldots, h_k\}$

C. $\sum_{i=1}^{k} h_i$

iii. Repeat (ii) for the case where you are allowed to move only one knight at a time.

(c) Consider a search problem in which the states are nodes along a path $n_0, n_1 \ldots n_m$, where n_0 is the start state, n_m is the goal state and for all i there is an action from n_i to n_{i+1} of cost 1. The cheapest cost from any node n_i to the goal is then $c(n_i) = m - i$. Let a heuristic function be defined as $h(n) = m - 2 \times \lceil \frac{i}{2} \rceil$, where $\lceil x \rceil = \text{ceiling}(x)$ is the smallest integer not less than x. *(8 marks)*

i. Is $h(n)$ admissible? Explain.

ii. Is $h(n)$ consistent? Explain.

3. Constraint Satisfaction Problems

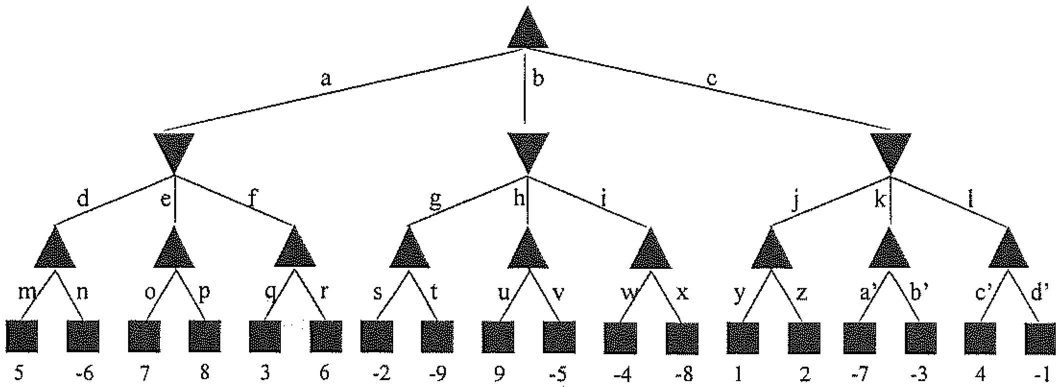
(a) Write down a brief and precise definition of the following. *(10 marks)*

i. Cycle cutset

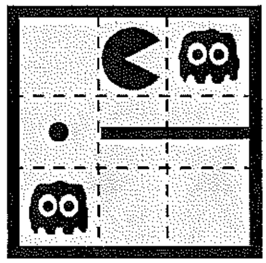
ii. Min-conflicts

4. Adversarial Search

(a) For the game tree shown below, which branches will be pruned by alpha-beta pruning? What is the minimax value at the root? *(8 marks)*



(b) Consider the maze below. Assume that Pac-man moves first, followed by the lower left ghost, then the top right ghost. Ghosts cannot change direction unless they are facing a wall. The possible actions are east, west, south, and north (not stop). Initially, they have no direction and can move to any adjacent square. Assume that Pac-man cannot stop.



The game is scored as follows

- -1 for each action Pac-man takes
- 10 for each food dot eaten
- -500 for losing (if Pac-man is eaten)
- 500 for winning (all food dots eaten)

Consider minimax ghosts, i.e., assume ghosts are minimizing nodes. Build the minimax tree of this game and draw Pac-man's optimal decision at the root of the tree. *(6 marks)*

(c) Consider the following lotteries

- $A : [0.8, \$4000; 0.2, \$0]$
- $B : [1.0, \$3000; 0.0, \$0]$
- $C : [0.2, \$4000; 0.8, \$0]$
- $D : [0.25, \$3000; 0.75, \$0]$

Show that the judgments $B \succ A$ and $C \succ D$ violate the axiom of substitutability. *(6 marks)*

5. Markov Decision Processes

(a) Bob notices value iteration converges more quickly with smaller γ and rather than using the true discount factor γ, he decides to use a discount factor of $\alpha\gamma$ with $0 < \alpha < 1$ when running value iteration. Mark each of the following that are guaranteed to be true and optionally provide a short explanation in the space provided. *(8 marks)*

- ☐ While Bob will not find the correct value, he could simply rescale the values he finds by $\frac{1-\gamma}{1-\alpha}$ to find the correct value.

- ☐ If the MDP's transition model is deterministic and the MDP has zero rewards everywhere, except for a single transition at the goal with a positive reward, then Bob will still find the optimal policy.

- ☐ If the MDP's transition model is deterministic, then Bob will still find the optimal policy.

- ☐ Bob's policy will tend to more heavily favor short-term rewards over long-term rewards compared to the optimal policy.

(b) Suppose that we define the utility of a state sequence to be the maximum reward obtained in any state in the sequence. Show that this utility function does not result in stationary preferences between state sequences. *(7 marks)*

(c) Consider the following transition and reward function for an MDP that has two states and two actions.

s	a	s'	$T(s,a,s')$	$R(s,a,s')$
A	0	A	0.5	2
A	0	B	0.5	-1
A	1	A	0.5	1
A	1	B	0.5	2
B	0	A	0.1	-2
B	0	B	0.9	-1
B	1	A	0.2	-3
B	1	B	0.8	-1

Let the discount factor γ be 1.0. Determine the following values using value iteration. *(10 marks)*

- $V_0(A), V_0(B)$
- $Q_1(A,0), Q_1(A,1), Q_1(B,0), Q_1(B,1)$
- $V_1(A), V_1(B)$
- $Q_2(A,0), Q_2(A,1), Q_2(B,0), Q_2(B,1)$
- $V_2(A), V_2(B)$

6. Reinforcement Learning

(a) Write down the main difference between planning in a known Markov Decision Process (MDP) and Reinforcement Learning (RL). *(2 marks)*

(b) What is active learning and how does is differ from passive learning? *(4 marks)*

(c) For Q-Learning to converge to the optimal Q-values ... *(4 marks)*

- ☐ It is necessary that every state-action pair is visited infinitely often.
- ☐ It is necessary that the learning rate α (the weight given to new samples) is decreased to 0 over time.
- ☐ It is necessary that the discount γ is less than 0.5.
- ☐ It is necessary that actions get chosen according to $\arg\max_a Q(s, a)$

(Mark all that apply)

END OF PAPER

Please be reminded to write your university number on the first page.

This page is intentionally left blank.
You may use it to draft your answers.

This page is intentionally left blank.
You may use it to draft your answers.