香 港 中 文 大 學

The Chinese University of Hong Kong

二0一二至一三年度 上學期科目考試

Course Examination 1st Term, 2012 – 2013

版權所有 不得翻印
Copyright Reserved

科目編號及名稱 Course Code & Title : CSCI3230 Fundamentals of Artificial Intelligence

時間 Time allowed : 2 小時 hours 00 分鐘 minutes

學號 Student I.D. No. : _____ 座號 Seat No. : _____

Answer Four out of Five questions
Total: 100 marks

1.

a) Translate the following three first-order sentences into English given that hate(x,y) means x hates y. [3 marks]

i. $\forall x \exists y\ hate(x,y)$.
ii. $\exists x \forall y\ hate(x,y)$.
iii. $\exists y \forall x\ hate(x,y)$.

b) Express each of the following statements (i to vi) as First Order Logic sentences using the symbols below (together with $\forall, \exists, \neg, \wedge, \vee, \Rightarrow$): [6 marks]

Symbol	Meaning
Student(X)	X is a student
Task(X)	X is a task
Likes(X,Y)	X likes Y
RC(X)	X requires creativity
RL(X)	X requires long working hours
Tough(X)	X is tough
Boring(X)	X is boring

i. Students like only tasks which are neither boring nor tough.
ii. *Programming* is a task which requires creativity.
iii. Any task requiring creativity requires long working hours.
iv. Any task requiring long working hours is either tough or boring or both.
v. *Antonio* does not like boring tasks.
vi. If *Antonio* likes *Programming*, then *Antonio* is not a student.

c) Convert your answers in (b) into **Conjunctive Normal Form** (CNF). [6 marks]

d) Prove statement vi from statements i to v using either **Resolution** approach or otherwise. [5 marks]

e) What are the differences between propositional logic, first-order logic and higher-order logic? [5 marks]

2.

a) Name and describe a multi-point stochastic search algorithm. [3 marks]

b) Alice and Bob are old friends. They enter the same university but seldom meet each other. One day Alice wants to invite Bob to have a lunch. Since Alice has lost Bob's contact, she asks her friends to pass the invitation message verbally to Bob.

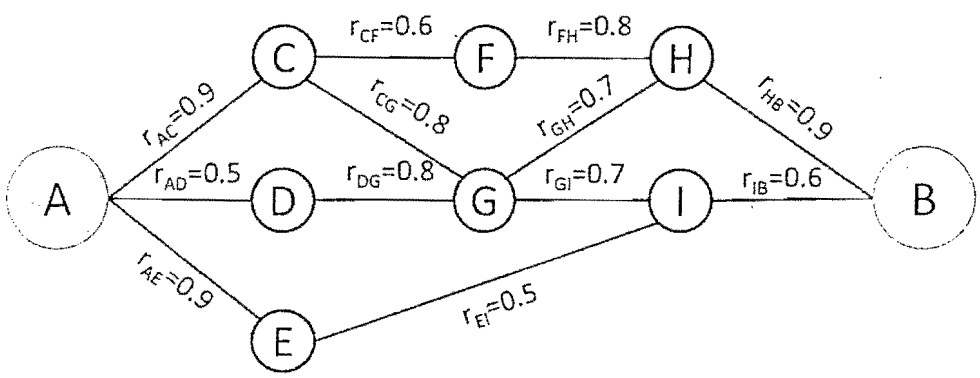
However, each person has a chance to lose a part of the message. The graph below shows the social network connecting Alice (A) and Bob (B). An edge represents a channel to pass the message. Each edge from person x to person y is assigned with a retention rate r_{xy}, where $0.5 \leq r_{xy} \leq 0.9$, which is the retained proportion of a message.

Considering the whole transmission, the total retention rate is defined as the product of the involved retention rates. For example, if the path is {A, D, G, I, B}, the overall retention rate is $r_{AD} \times r_{DG} \times r_{GI} \times r_{IB}$. Under this model, the goal is to find the path achieving the largest overall retention rate.

i. Since the total message retained is multiplicative, we want to transform it to additive path cost. As a result, the goal is to find a path with the **LEAST** cost in the transformed graph. Propose a transformation function and show the transformed retention rate for each edge (Hint: A log function). Round your calculation to three significant figures. [2 marks]

ii. Using the function proposed in (i), design a heuristic function for your algorithm and prove that it is admissible. (Hint: A function of number of hops) [4 marks]

iii. Perform A* search to find the best route to pass the message using your own heuristic function. Draw the search trees. [8 marks]



c) Describe briefly the four major evaluation criteria for search strategies and algorithms. [6 marks]

d) One of the environment properties of the tic-tac-toe task environment is strategic-deterministic. Name two more properties. [2 marks]

3.

a) Consider the following puzzle: A 3×3 grid with the label $X_1, X_2, \ldots, X_9$. Digits from 1 to 9 are allocated to the grid (Figure 1). Four kinds of actions are allowed:

- L_i: rotate the i-th row to the left
- R_i: rotate the i-th row to the right
- U_i: rotate the i-th column upward
- D_i: rotate the i-th column downward

Examples of possible movements are shown below (Figures 2, 3 and 4). The objective of this puzzle is to restore the puzzle to the arrangement in Figure 2. Formulate the puzzle as a search problem (state space, state transitions and goal state(s)). [6 marks]

X_1	X_2	X_3
X_4	X_5	X_6
X_7	X_8	X_9

Figure 1

1	2	3
4	5	6
7	8	9

Figure 2

L_1 ⇄ R_1

3	1	2
4	5	6
7	8	9

Figure 3

D_2 ⇄ U_2

3	8	2
4	1	6
7	5	9

Figure 4

b) Use *any heuristic* you like to solve the following instance. [3 marks]

6	5	3
7	4	8
9	1	2

c) The following is the pseudocode of Minimax algorithm using recursive negation. Modify the pseudocode into **alpha-beta pruning**. [6 marks]

```
function minimax(node, depth)
    if node is a terminal node or depth == 0
        return the heuristic value of node
    else
        best = -∞
    for child in node          // evaluation is identical for both players
         best = max(best, -minimax(child, depth-1))
    return best
```

d) In non-deterministic games, describe briefly how you would apply Minimax algorithm. [4 marks]

e) In real problems, the search spaces are usually very large. Suggest **TWO** methods to handle the huge spaces. [4 marks]

f) Briefly explain the **Ockham's Razor** principle. [2 marks]

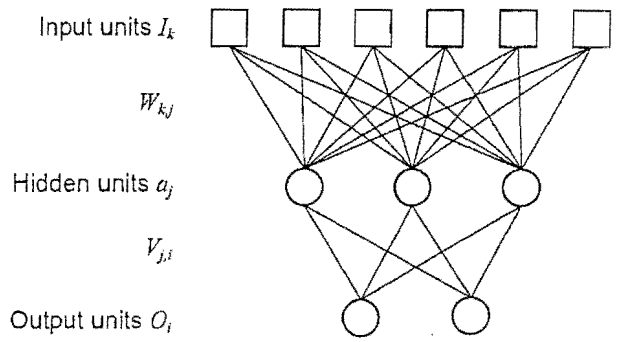
4.

a) Can a single neuron perceptron with two inputs be used to simulate "XOR Gate" logic computation? Explain why and illustrate your answer with an appropriate figure. [3 marks]

b) What are Single-Layer Feed-Forward Network and Multi-Layer Feed-Forward Network? Illustrate your answer with two appropriate figures. [3 marks]

c) For the sigmoid function $f(z)=\frac{1}{1+e^{-bz}}$, show that its first derivative $f'(z) = bf(z)(1-f(z))$. [3 marks]

d) Derive the **TWO** updating rules ($\frac{\partial E}{\partial V_{j,i}}$ and $\frac{\partial E}{\partial W_{k,j}}$) of weights and biases in back-propagation algorithm for Multi-Layer Feed-Forward Network (MLFFN). The activation function in the output units is the above sigmoid function, i.e., $O_i = f(u_i)$, where $u_i = \sum_j a_j \times V_{j,i} + B_{O_i}$. The activation function in the hidden units is the above sigmoid function, i.e., $a_j = f(s_j)$, where $s_j = \sum_k I_k \times W_{k,j} + B_{a_j}$. The error function is the sum of squared error, i.e., $E = \frac{1}{2}\sum_i (O_i - T_i)^2$ where T_i is the target value. The notations and the diagram of MLFFN are shown below: [10 marks]



e) When training a neural network, a beginner observes that when he increases the learning rate greatly, the total error increases with the training epochs. What is the problem? What is your advice to him? Explain briefly. [2 marks]

f) When training a neural network, a beginner observes that when he initializes all the weights to be ZERO, the total error converges slowly. What is the problem? What is your advice to him? Explain briefly. [2 marks]

g) When training a neural network, a beginner observes that when he increases the number of training epochs, the training accuracy will increase, but the validation accuracy decreases accordingly. What is the problem? What is your advice to him? Explain briefly. [2 marks]

5.

DNA Sequence	Attributes						Coding Region?
	Pos1	Pos2	Pos3	Pos4	Pos5	Pos6	
S_1	C	T	A	A	G	T	No
S_2	C	G	A	C	G	T	Yes
S_3	C	G	A	G	G	C	No
S_4	C	G	A	G	G	T	Yes
S_5	C	G	A	A	G	C	No
S_6	C	G	A	A	G	T	Yes
S_7	C	T	A	T	G	T	No
S_8	C	G	A	A	G	C	No
S_9	C	G	A	T	G	T	Yes
S_{10}	A	G	A	G	G	T	No
S_{11}	C	T	A	C	G	C	No

Table 1: Training Dataset

DNA Sequence	Attributes						Coding Region?
	Pos1	Pos2	Pos3	Pos4	Pos5	Pos6	
S_{12}	C	T	A	A	G	C	No
S_{13}	C	G	A	C	G	T	Yes
S_{14}	A	G	A	G	G	T	Yes
S_{15}	A	G	A	T	G	T	No
S_{16}	C	T	A	A	G	T	Yes

Table 2: Validation Dataset

a) We need to classify whether a segment of DNA is a (protein) coding region based on the data in Table 1. There are altogether 11 training samples (Table 1) and 5 validation samples (Table 2). Please use Information Gain (IG) to learn **a decision tree** based on the attributes in the training dataset and draw a diagram of the decision tree. **Show the detailed calculations on choosing the root attribute only.** You may find the following logarithm table useful. [10 marks]

x	1/2	1/3	2/3	1/4	3/4	3/7	4/7
$\log_2(x)$	-1	-1.5850	-0.5850	-2	-0.4150	-1.2224	-0.8074
x	1/5	2/5	3/5	4/5	4/11	7/11	
$\log_2(x)$	-2.3219	-1.3219	-0.7370	-0.3219	-1.4594	-0.6521	

b) Evaluate the decision tree using the validation dataset. Construct a **Confusion Matrix** to show the results, using "Yes" as positive and "No" as negative. Label the meaning of the rows and the columns of your matrix. [4 marks]

c) Based on your **Confusion Matrix**, calculate each of the following performance metrics: $Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$, $Precision = \frac{TP}{TP+FP}$, $Recall = \frac{TP}{TP+FN}$, $F\_measure = \frac{1}{\left(\frac{1}{Recall}+\frac{1}{Precision}\right)/2}$. [4 marks]

d) What is cross-validation? What is over-fitting? How does cross-validation prevent against over-fitting? [5 marks]

e) Name one advantage and one disadvantage of the training-evaluation method (in (a) and (b)) over cross-validation. [2 marks]

____ 全卷完 ____

END