

NAME:

**SOLUTIONS**

Signature:

Student Nr.



**EASTERN MEDITERRANEAN UNIVERSITY  
COMPUTER ENGINEERING DEPARTMENT  
CMPE 535 - Knowledge Engineering**

Spring 2006-07

**Midterm Exam**

Q	1	2	3	4	5	6	7	8	9	10	11	12	total
pts	30	20	4	4	4	4	4	4, 2	4	4	4, 4	2, 2, 4	100

1) [30p] Consider the following kids-outside-play decision data of the parents.

No.	Outlook	Temperature	Humidity	Windy	Play
1	sunny	hot	high	false	no
2	sunny	hot	high	true	no
3	overcast	hot	high	false	yes
4	rain	mild	high	false	yes
5	rain	cool	normal	false	yes
6	rain	cool	normal	true	no
7	overcast	cool	normal	true	yes
8	sunny	mild	high	false	no
9	sunny	cool	normal	false	yes
10	rain	mild	normal	false	yes
11	sunny	mild	normal	true	yes
12	overcast	mild	high	true	yes
13	overcast	hot	normal	false	yes
14	rain	mild	high	true	no

a) Is this data set structured? If it is structured then specify its structure. ....  
it is structured in table format.

b) Does it contain explicit information? Give an example of information emphasizing the differences between data, information and knowledge .....  
Information is structured data. Knowledge is condensed form of information. Data is structured, and the titles of each column provides information to comprehend the meaning of each row. For example first row says if **outlook is sunny, and temperature is hot and humidity is high, and not windy, then playing outside is not allowed to the kid.** This is an information.

c) Write the arguments of the domain and the solution space of this data set.  
Outlook, Temperature, Humidity, Wind  
Domain Space arguments: .....  
Play = Kids allowed to play at outside.  
Solution Space arguments: .....  
Outlook, Temperature, Humidity, Wind, Play

d) Inductive Decision Tree (ID3) method finds out the simplest set of rules is called algorithm based on the following index.

$$\begin{aligned} \text{inf}(C, a) &= -p \cdot \log_2(p) - (1 - p) \log_2(1 - p), \\ \text{inf}(C, \sim a) &= -q \cdot \log_2(q) - (1 - q) \log_2(1 - q), \\ \text{INF}(C, a) &= p \text{inf}(C, a) + (1 - p) \text{inf}(C, \sim a), \end{aligned}$$

where C is the class of passes, a is the binary test attribute, ~a is complement of attribute, p is the pass rate in the partition satisfied by a, q is the fail rate in the partition satisfied by ~a.

What is the conceptual criteria of ID3 algorithm in ranking the significance of each argument?

INF(C, a) is a measure of loss of information in the representation. Simplicity being the same, minimum loss of information gives the best representation.

.....

2) [20p] Assume that the following data set describes the diagnosis of the choulepia illness.

1	AL	GN	BN	LH	CP
2	AL	GH	BN	LL	CP
3	AL	GN	BN	LL	CN
4	AL	GN	BH	LL	CN
5	AN	GN	BN	LL	CN
6	AN	GN	BH	LL	CN
7	AN	GN	BN	LH	CP
8	AH	GN	BN	LH	CP
9	AH	GH	BN	LL	CP
10	AH	GN	BH	LL	CN

AL: Arter-pressure low;  
 AN: Arteric-pressure normal;  
 AH: Arter-pressure high;  
 GN: Glucose-normal;  
 GL: Glucose-low;  
 LL: Leucocyte count low;  
 LN: Leucocyte count normal,  
 LH: Leucocyte count high;  
 CN: Choulepia-negative;  
 CP: Choulepia-positive.

a) Apply the ID3 algorithm to compare the significance of the nodes LL/~LL and AL/~AL to decide which of them must come first in the decision tree. Complete the following table for your calculations

case under test: (a)	AL	LL
nr. of pass in the partition satisfied by a	2	2
nr. of fail in the partition satisfied by a	2	5
nr. of fail in the partition satisfied by ~a	3	0
nr. of pass in complement partition	3	3
(p) pass rate for the test	0.5	2/7
(q) fail rate for complement of test	0.5	0
$p \log_2(p)$	$0.5 \times -1 = -0.5$	$-1.81 \times 2/7 = -0.516$
$(1-p) \log_2(1-p)$	$0.5 \times -1 = -0.5$	$-0.49 \times 5/7 = -0.35$
$q \log_2(q)$	$0.5 \times -1 = -0.5$	0
$(1-q) \log_2(1-q)$	$0.5 \times -1 = -0.5$	0
$\text{inf}(a) = -p \log_2(p) - (1-p) \log_2(1-p)$	$0.5 + 0.5 = 1$	$0.516 + 0.35 = 0.87$
$\text{inf}(\sim a) = -q \log_2(q) - (1-q) \log_2(1-q)$	$0.5 + 0.5 = 1$	0
$\text{INF}(a) = p \text{inf}(a) + (1-p) \text{inf}(\sim a)$	1	0.25

Which one of the two arguments discriminate better?

Table of  $(a/b) \times \log_2(a/b)$   
 for the integers  $a=1, \dots, 10$  and  $b=1, \dots, 10$ .

b ↓	a									
	1	2	3	4	5	6	7	8	9	10
1	0.00									
2	-0.50	0.00								
3	-0.53	-0.39	0.00							
4	-0.50	-0.50	-0.31	0.00						
5	-0.46	-0.53	-0.44	-0.26	0.00					
6	-0.43	-0.53	-0.50	-0.39	-0.22	0.00				
7	-0.40	-0.52	-0.52	-0.46	-0.35	-0.19	0.00			
8	-0.38	-0.50	-0.53	-0.50	-0.42	-0.31	-0.17	0.00		
9	-0.35	-0.48	-0.53	-0.52	-0.47	-0.39	-0.28	-0.15	0.00	
10	-0.33	-0.46	-0.52	-0.53	-0.50	-0.44	-0.36	-0.26	-0.14	0.00

$x \rightarrow +0 \text{ Limit } x \log_2 x \rightarrow 0.$

b) Can you apply the Karnough Map method to obtain the minimal decision tree of the described diagnosis problem? Why?

K-map is useful only with binary attributes. In this case, some attributes are multi-valued, and we need a pre-processing to convert the problem to binary before applying K-map on this problem.

3) a) What are the two compromising factor in a good data representation for knowledge engineering.

Simplicity and unambiguity (=minimum loss of information)  
 ..... and .....

b) Is data representation of problem-2 successful in both respects? .....  
 yes, it provides simplicity for minimal loss of information

c) What other alternative representations are suitable for the same data? .....  
 Lists, sets, frames

4) Explain the basic difference between the first order logic systems and the higher order logic systems considering their rule formats and inference methods.

First order logic cannot infer to *socrates is mortal* from *all human are mortal*, and *socrates is human*,. First order logic only verifies the truth tables. Higher order logic can infer through backward and forward chaining.

.....  
 .....

5) a) How can you test the generalization and overfitting?

By using a verification or test data set, which is independent to the training data set

6) Complete the list of major issues of the Knowledge Systems:

- |                         |                         |
|-------------------------|-------------------------|
| 1- representation,      | 2- inference,           |
| 3- learning.....,       | 4- generalization....., |
| 5- interaction,         | 6- explanation.....,    |
| 7- validation....., and | 8- adaptation           |

7) Explain the contents of explanation to the question “why” in the symbolic logic systems on the example of car diagnosis system:

(Temperature\_over\_120 and gauge\_is\_ok ) or Breaks-slow → stop\_the\_car

Why tracks explanation in backward direction. Stop the car because either breaks are slow, or both temperature is over 120, and also gauge is ok.

8) a) In the following example indicate the propositional symbols, and the connectives

$R = ( P \vee Q ) \wedge P \wedge Q$  P,Q, and R are propositional symbols,  $\vee$  and  $\wedge$  are connectives

b) Consider the following two logic expressions. What is the type of the logic system?  
 (predicate or propositional?)

$\forall x, \quad x \in \text{Human} \rightarrow x \in \text{Mortal} .$   
**Socrates  $\in$  Human .**

If it infers Socrates is Mortal, then it is a predicate logic.

9) a) The CART (Classification and Regression Tree) method constructs binary trees for regression by splitting the arguments in binary ranges with minimal loss of information. For example, the income variable x that changes within the interval  $x \in [20\ 000, 100\ 000]$  is split in two ranges by a condition if  $x > 40000$  then  $x_0=1$ ; otherwise  $x_0=0$ ;

Write the most important two effects of splitting a variable considering the basic requirements of data representation.

Splitting simplifies representation. but Information loss is inevitable, this method minimizes the information loss.

.....

10) Shemata are more general problem representation format than semantic web. Why?  
 Give an example which cannot be represented by a semantic web, but easily represented by shemata.  
 Transition of states (FSM) and probabilities of state transitions (Markov Chain Models) cannot be shown by semantic web.

11) Explain the matching process in CLIPS and in PROLOG. Give one example and explain how this example is executed in the both systems.  
 Clips uses forward chaining, Prolog uses backward chaining.

12) a) What is the meaning of  $p(A) = 0.5$ , ..... event A occurs at half of all events.  
 probability of a fact A is  $p(A) = 0.5$ , .....

b) Apply Bayes rule (i.e.,  $p(A|C) = p(C|A) \cdot p(A) / p(C)$ ) to infer the probability  $p(A|C)$  for the following rule  
 if ( C1 and C2 ) then A  
 given that the prior probabilities  $p(C1) = 1/2$  .  $p(C2) = 1/3$ ,  $p(A) = 1/12$ , and  $p(C1, C2 | A) = 1/4$  ,  
 assuming that C1 and C2 are independent events.

$$p(A|C) = \dots\dots\dots = 1/4 \times ( 1/12 / ( 1/2 \times 1/3 ) ) = 1/8$$

What is the main disadvantage of the probabilistic inference methods?  
 If conditions are not independent, then bayes rule does not apply.  
 .....

c) What is the interval of the following uncertainty measures:

plausibility  $\in [ \dots\dots\dots , \dots\dots\dots ] ; 0 \dots 1$

belief  $\in [ \dots\dots\dots , \dots\dots\dots ] ; 0 \dots 1$

certainty factor  $\in [ \dots\dots\dots , \dots\dots\dots ] ; -1 \dots 1$

possibility  $\in [ \dots\dots\dots , \dots\dots\dots ] ; 0 \dots 1$