Knowledge-based systems in bioinformatics – 1MB602

Exam 2007-04-17

Contact under exam: Robin Andersson 018-471 66 86

Grading:
- 3: ≥ 20p
- 4: ≥ 27p
- 5: ≥ 36p

Max: 45p

General instructions:
1. Keep it short!
2. If anything is unclear, clearly state your assumptions.

Good luck!

1) (2+2p)

Consider the procedure `order` that orders a list of elements according to a list of positions. The second list specifies the new positions of the elements in the original list. Both lists are assumed to have equal length.

Example:

```
(order '(a b c d e) '(1 2 3 4 5))
; Value 1: (a b c d e)
(order '(a b c d e) '(4 1 2 3 5))
; Value 2: (b c d a e)
(order '(a b c d e) '(5 4 3 2 1))
; Value 3: (e d c b a)
(order '(a b c d e) '(1 5 2 4 3))
; Value 4: (a c e d b)
```

(a) (2p) Implement `order` using a recursive process.
(b) (2p) Implement `order` using an iterative process.

2) (6p)

Define a **max-priority queue** using **message-driven programming**, i.e. with a dispatch procedure and local state. A priority queue is a data structure for maintaining a set *S* of elements, each associated with a value called a key. A max-priority queue supports the following three operations:

- INSERT(*S*,x,k): insert the element *x* with key *k* into the set *S*.
- MAXIMUM(*S*): returns the element of *S* with the largest key.
- EXTRACT-MAX(*S*): removes and returns the element of *S* with the largest key.

Your code should support the following scenario:

```
(define S (make-max-priority-queue))
(insert S ‘a 3)
(insert S ‘b 4)
(insert S ‘c 2)
(insert S ‘d 1)
(maximum S)
;Value 1: b
(extract-max S)
;Value 2: b
(maximum S)
;Value 3: a
```

3) (3+2p)

(a) (3p) Decide, using truth tables, whether each of the following logical implications in propositional logic is satisfiable, contingent, valid (tautology), and/or contradictory.

1. $\models (A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$
2. $A \rightarrow \neg B, B \vee C, C \rightarrow A \models A \vee C$
3. $(A \rightarrow B) \wedge (C \rightarrow D) \models (A \vee C) \rightarrow (B \vee D)$

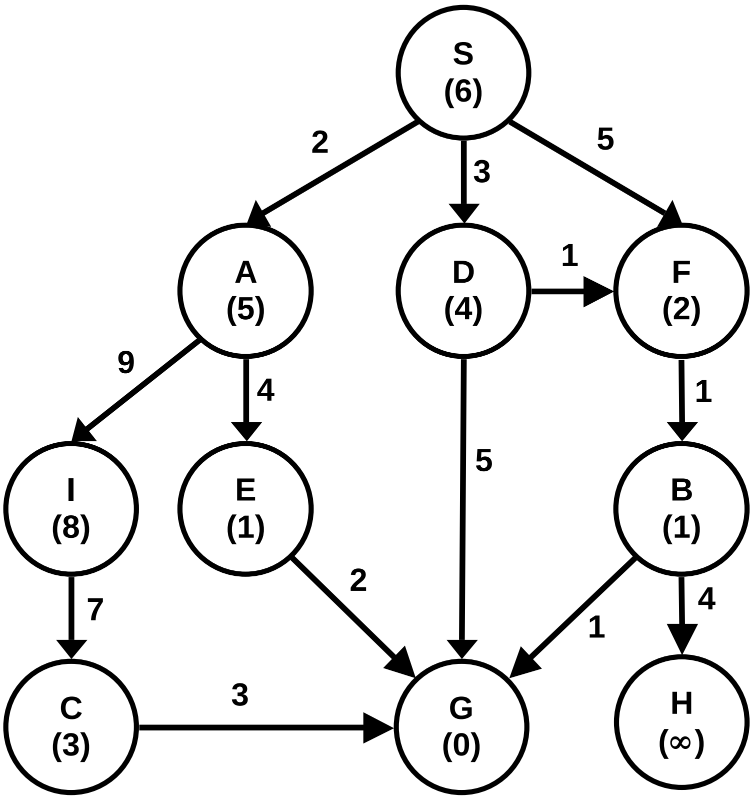
(b) (2p) Provide a proof of validity in natural deduction for one of the logical implications you find valid (inference rules are listed on the last page).

4) (4p)

Show that the connectives ($\neg,\wedge,\vee,\leftrightarrow$) in propositional logic can be written using only the connective $\rightarrow$ (implication) and the sentence $\perp$ (the sentence which is always false). Hint: use truth tables.

5) (6p)

Consider the following graph where the numbers by the edges define the cost of following that edge and the numbers in parentheses in the nodes give the estimated cost of going from that node to the goal. The initial state is S and the goal state is G. Using Depth First Search (DFS), Breadth First Search (BFS), Iterative Deepening Search (IDS) using DFS, and A* Search show which **path** each algorithm finds and in which **order** the nodes are visited. All assumptions must be clearly stated in the solutions.



6) (3+1p)

(a) (3p) What is meant by the following terms considering search algorithms?

- Completeness
- Soundness
- Optimality

(b) (1p) Under what condition is SMA* search optimal?

7) (2p)

When would the genetic algorithm benefit from using a generational approach? When is a steady state approach more beneficial?

8) (6p)

Consider the following decision table describing factors affecting sunburn (hair color, height, weight, and whether lotion is used or not).

Hair	Height	Weight	Lotion	Decision
Blonde	Average	Light	No	Sunburned
Blonde	Tall	Average	Yes	None
Brown	Short	Average	Yes	None
Blonde	Short	Average	No	Sunburned
Red	Average	Heavy	No	Sunburned
Brown	Tall	Heavy	No	None
Brown	Average	Heavy	No	None
Blonde	Short	Light	Yes	None

Build a decision tree from the decision table using the gain criterion for feature selection.

Hint: log2(a/b) = log2(a) – log2(b), log2(1) = 0, log2(2) = 1, log2(3) ≈ 1.58, log2(4) = 2, log2(5) ≈ 2.32, log2(6) ≈ 2.58, log2(7) ≈ 2.81, log2(8) = 3

9) (4+4p)

Given a set of data measurements, the task of clustering procedures is to find the best grouping of data into a specified number of clusters that yields the minimum distance within clusters and the maximum distance between clusters.

(**a**) (4p) Formulate the clustering problem as a search problem. Specify representation, operators, and a strategy for searching the state space. (max 0.5 page)

(b) (4p) Provide a strategy for solving the problem using a genetic algorithm. Specify representation, selection operator(s), mutation operator(s), crossover, and a strategy for generating new generations of solutions. (max 0.5 page)

Inference rules:

$$\frac{\alpha,\ \neg\alpha}{\bot}\ (\bot I)$$

$$\frac{\boxed{\begin{array}{l}\alpha\\ \bot\end{array}}}{\neg\alpha}\ (\neg I) \qquad \frac{\boxed{\begin{array}{l}\neg\alpha\\ \bot\end{array}}}{\alpha}\ (\neg E) \qquad \frac{\alpha \rightarrow \beta,\ \beta \rightarrow \alpha}{\alpha \leftrightarrow \beta}\ (\leftrightarrow I)$$

$$\frac{\alpha_1,\ \alpha_2,\ \ldots,\ \alpha_n}{\alpha_1 \wedge \alpha_2\ \wedge \ldots \wedge \alpha_n}\ (\wedge I) \qquad \frac{\alpha_1 \wedge \alpha_2\ \wedge \ldots \wedge \alpha_n}{\alpha_i}\ (\wedge E) \qquad \frac{\alpha \leftrightarrow \beta}{\alpha \rightarrow \beta}\ (\leftrightarrow E)$$

$$\frac{\alpha_i}{\alpha_1 \vee \alpha_2\ \vee \ldots \vee \alpha_n}\ (\vee I) \qquad \frac{\alpha \vee \beta,\ \alpha \rightarrow \chi,\ \beta \rightarrow \chi}{\chi}\ (\vee E)$$

$$\frac{\boxed{\begin{array}{l}\alpha\\ \beta\end{array}}}{\alpha \rightarrow \beta}\ (\rightarrow I) \qquad \frac{\alpha \rightarrow \beta,\ \ \alpha}{\beta}\ (\rightarrow E)$$