FALL 2004
COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN – MADISON
PH.D. QUALIFYING EXAMINATION

Artificial Intelligence

Monday, September 20, 2004
3:00 – 7:00 p.m.
Room 144 Chemistry Hall

GENERAL INSTRUCTIONS:

1. Answer each question in a separate book.

2. Indicate on the cover of *each* book the area of the exam, your code number, and the question answered in that book. On *one* of your books, list the numbers of *all* the questions answered. *Do not write your name on any answer book.*

3. Return all answer books in the folder provided. Additional answer books are available if needed.

SPECIFIC INSTRUCTIONS:

Answer:

- both (2) questions in the section labeled B760 or B766, corresponding to your chosen focus area, *and*

- any two (2) additional question in the sections B731, B760, B766, and B776, where these two questions need *not* come from the same section, *and*

- both (2) questions in the section labeled A7xx that corresponds to your focus area.

Hence, you are to answer a total of *six* (6) questions.

POLICY ON MISPRINTS AND AMBIGUITIES:

The Exam Committee tries to proofread the exam as carefully as possible. Nevertheless, the exam sometimes contains misprints and ambiguities. If you are convinced that a problem has been stated incorrectly, mention this to the proctor. If necessary, the proctor can contact a representative of the area to resolve problems during the *first hour* of the exam. In any case, you should indicate your interpretation of the problem in your written answer. Your interpretation should be such that the problem is nontrivial.

Answer both (2) of the questions in the section labeled B7xx that corresponds to your chosen focus area. Also answer any two (2) additional questions in any of the other sections (these two questions need NOT occur in the same section).

B731 – ADVANCED AI BASIC QUESTIONS

B731-1. Suppose you are given a knowledge base of propositional definite clauses. Each clause has a real number between 0 and 1 attached. This number is the probability that the consequent (head) of the clause is true given that the antecedent (body) of the clause is true. A user will make queries to the knowledge base that also are definite clauses. The correct answer to a query is the probability that the head of the query is true given that the body is true. For illustration, see the knowledge base and query below.

(a) One approach that has been tried for answering such queries is the following. Add to the knowledge base the literals in the body of the clause, and see if resolution can be used to derive the head of the query. If so, multiply the probabilities of the clauses used in the resolution derivation; the result is the probability that the head of the query is true. If no derivation is found, the probability is 0. Discuss <u>two</u> significant problems with this approach.

(b) Describe one alternative to the approach in (a) that addresses these problems. If necessary, you may assume additional information (e.g., additional probabilities). Please explicitly describe any additional information you need.

Example Query: measles ← clammy, red_dots.

Example Knowledge Base with Probabilities:

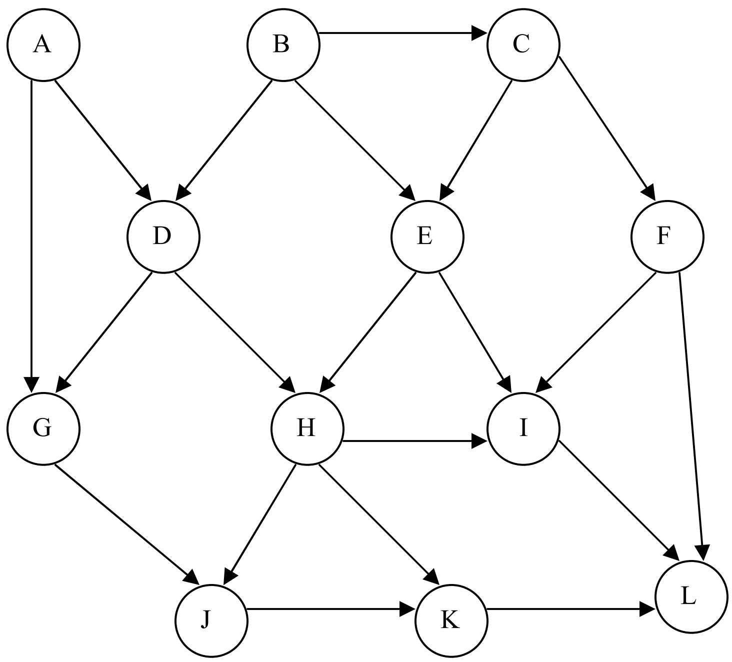
0.9	fever ← measles
0.8	fever ← flu
0.9	red_dots ← measles
0.01	measles ← fever
0.1	flu ← fever
0.5	measles ← red_dots
0.6	fever ← clammy

B731-2. Consider the Bayesian network below. Copy the network into your answer book.

(a) Show the edges you need to add for moralization (draw them on your graph as thick lines).

(b) Show any minimal set of edges you need to add to triangulate the resulting graph (draw the edges as dashed lines).

(c) Show a junction tree that results from your triangulated graph.



B760 – MACHINE LEARNING BASIC QUESTIONS

B760-1. Imagine you are tasked with creating a richly interconnected relational database describing TV shows, as well as data mining it to help choose which shows to include in the next TV season. This database would need to record for each show such information as the actors in the show, the network presenting it, the length of the show, the shows in the time slots immediately before and after it (on any channel), the shows on at the same time on other channels, the average number of viewers each week, etc. Assume you choose to predict whether or not the average number of weekly viewers will exceed 10 million people.

(a) Which of the above-mentioned attributes would be the hardest to usefully represent for a decision-tree learner, such as Quinlan's ID3 algorithm? Justify your answer.

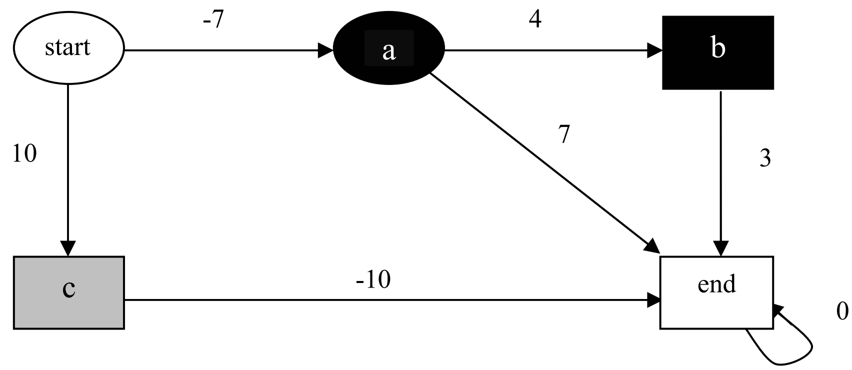
(b) If you wanted to represent the attribute "*show on immediately before on the same channel*" for a support-vector machine, how would you go about doing so?

(c) Assume you wish to apply a *k-nearest neighbor* algorithm to this task. Describe how you might define a distance function comparing the actors in two shows.

(d) Next, imagine that you choose to apply Quinlan's FOIL algorithm to this task. How would a real-valued attribute such as *show length* be handled?

(e) Finally, imagine you wish to employ *explanation-based learning* to this dataset. Present, using first-order predicate calculus, and explain <u>one</u> item in the *domain theory* you would use.

B760-2. Consider the *deterministic reinforcement environment* drawn below. The numbers on the arcs indicate the immediate rewards, and the shape and shading of the nodes represent the properties of the states. Let the discount rate $\gamma = 0.9$.



(a) Using a *Q table* where all entries are initialized to zero, show the state of this table *after* each of the following steps. You can represent actions using the following notation: **start → c, a → b, a → end,** etc. It is fine to report only the table entries that *change* after each action is performed. Be sure to explain your work.

i. the learner is placed (by its human trainer) in *start* and chooses to perform an *exploration* action

ii. the learned is returned (again by its human trainer) to *start* and follows its current policy

(b) Assume that this time you choose to use a perceptron (i.e., no hidden units) to represent the *Q* function; initialize all free parameters to zero and use a learning rate of 0.5 for the perceptron. Show how this could be done and concretely illustrate your solution by illustrating what happens after the learner chooses to go from state *a* to state *b* (i.e., assume that this is the *first* action executed after the perceptron is initialized).

B766 – COMPUTER VISION BASIC QUESTIONS

B766-1. Stereo Vision

(a) Describe the process of rectification that is commonly performed as part of a stereo algorithm, including what is its purpose.

(b) What happens in rectification if the focal lengths of the two cameras are not equal?

(c) What parameters affect the accuracy of the depth estimate to a scene point **P** at actual depth z_P assuming the only source of noise is the localization of corresponding points in two images from two pinhole cameras? Give the parameters and how (qualitatively) varying each will affect the accuracy of the depth estimate.

(d) What is the "ordering" constraint (also known as the "monotonicity" constraint) that is sometimes used in solving the correspondence problem? Give one major advantage and one major disadvantage of using this constraint in a stereo system.

B766-2. Two-Dimensional Object Recognition

(a) Describe the main steps of the geometric hashing algorithm for recognizing two-dimensional objects at arbitrary positions and orientations in an image.

(b) How many feature point correspondences are necessary for the geometric hashing algorithm, and what is the most general class of two-dimensional transformations that can be recognized?

(c) Comment on how well geometric hashing can cope with (i) partially occluded objects and (ii) errors in feature point localization.

B776 – BIOINFORMATICS BASIC QUESTIONS

B776-1.

(a) Suppose we were using a hierarchical clustering method to arrive at a partitional (i.e., non-hierarchical) clustering with **two** clusters. Given the following **distance** matrix, would *single-link*, *complete-link* and *average-link* clustering produce the same two clusters? Show your work to justify your answer.

	A	B	C	D	E
A	0	4	8	5	2
B		0	5	4	1
C			0	7	3
D				0	6
E					0

(b) A notable limitation of standard hierarchical clustering methods is that they are greedy algorithms. Describe how you might modify standard hierarchical clustering so that it employed a more extensive search process.

B776-2. Suppose we have trained a stochastic context free grammar (SCFG) to model a certain class of subsequences (e.g., a particular type of RNA genes) and we now want to use the SCFG to recognize instances of this class that are embedded in longer sequences (e.g. whole genomes).

(a) Briefly describe how you would use the CYK algorithm with this SCFG to identify instances of the class in longer sequences.

(b) What is the time and space complexity of the CYK algorithm? Informally justify your answer.

(c) Suppose that we need a faster approach to identifying class instances in sequences (the sequences being processed might be very long). Describe how we might modify the CYK approach to handle this situation.

(d) List <u>one</u> significant disadvantage of your proposed approach.

Answer both (2) of the questions in the section labeled A7xx that corresponds to your chosen focus area.

A760 – MACHINE LEARNING ADVANCED QUESTIONS

A760-1. Consider the task of creating a software assistant that learns to help its user make travel plans. Describe how each of the following machine learning approaches might be used in such a setting. Define each approach and provide one concrete example of how it could be used for this task or argue that the approach is inapplicable to creating a travel-planning assistant that learns.

(a) Active learning

(b) Collective inference

(c) Theory refinement

(d) Hierarchical reinforcement learning

A760-2. Consider the following relational database and the corresponding probabilistic relational model (PRM) below it.

(a) Show all conditional probability tables (CPT) of the PRM after parameters are estimated from the given database. Assume that pseudocounts = 1 are used for all parameters You may assume that *Grade* ranges over only 'a' and 'b,' while all other non-key fields range over only 'l' (low) and 'h' (high).

(b) Describe <u>one</u> significant advantage and <u>one</u> significant disadvantage of PRM learning compared with inductive logic programming (ILP).

Student

Key	Intelligence	Rank
Sam	l	l
Terry	h	h

Course

Key	Professor	Rating	Difficulty
cs1	Sue	h	h
cs2	David	h	l

Professor

Key	Popularity	Ability
Sue	h	h
David	l	l

Registration

Key	Course	Student	Satis-faction	Grade
r1	cs1	Sam	h	a
r2	cs2	Sam	l	b
r3	cs1	Terry	h	b