FALL 2003
COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN – MADISON
PH.D. QUALIFYING EXAMINATION

Artificial Intelligence

Monday, September 15, 2003
3:00 – 7:00 p.m.
Room 2540 Engineering 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.

(a) List two major advantages of inductive logic programming (ILP) over Naïve Bayes for mining a database containing multiple relational tables.

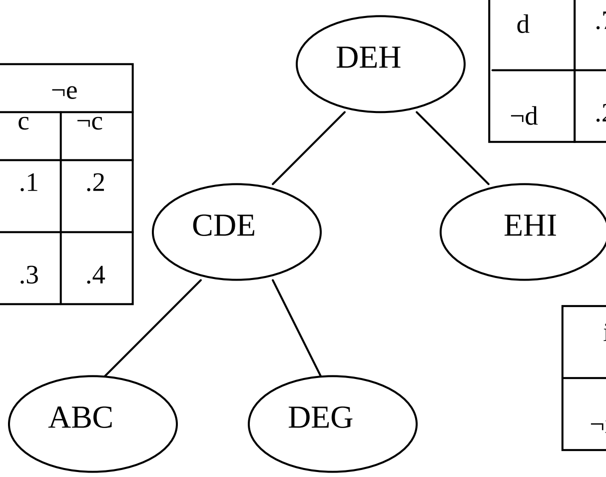
(b) Support your answer to (a) using a concrete example of a relational database with at least three relational tables. First describe/draw the schema for your relational database (tables and fields), then describe how ILP can be applied to this database, and finally describe the difficulties in applying Naïve Bayes to this database.

(c) Suppose you are committed to using Naïve Bayes to mine the database in (b). How will you proceed?

B731-2. Given the following junction tree, incorporate the evidence ¬e (i.e., E = false), and perform the upward and downward passes. Show all the tables at the end of the procedure (after both the upward and downward passes).

	e		¬e	
	h	¬h	h	¬h
d	.7	.3	.4	.6
¬d	.2	.8	.9	.1

	e		¬e	
	c	¬c	c	¬c
d	.1	.2	.1	.2
¬d	.3	.2	.3	.4



	e		¬e	
	h	¬h	h	¬h
i	.7	.2	.1	.6
¬i	.3	.8	.9	.4

	b		¬b	
	c	¬c	c	¬c
a	..2	.1	.05	.1
¬a	.05	.2	.2	.1

	e		¬e	
	g	¬g	g	¬g
d	.9	.2	.2	.1
¬d	.1	.5	.6	.1

B760 – MACHINE LEARNING BASIC QUESTIONS

B760-1. Both logic-based and probability-based approaches to machine learning exist.

(a) How might you alter the basic FOIL algorithm so that each time the user invokes it on the *same* training set it independently and stochastically learns a set of up to k conjunctive rules?

(b) Assume that you run the configuration you created for Part (a) n times, setting $k = 1$ in each run. How might you use the Naïve Bayes algorithm to try to improve the resulting set of rules? (Assume that you have pre-partitioned your data into disjoint training, tuning, and testing sets.)

(c) Describe one (1) key advantage and one (1) key weakness of the logic-probability hybrid algorithm you devised in parts (a) and (b).

B760-2. Imagine that you work at a video-game company and your task is to devise a way to employ machine learning to create artificially intelligent characters inside your company's newest video game. (To be specific, assume that in this game the human player plays the role of a cat that has to catch mice in a barn. Your job is to create smart *mice* via machine learning.)

(a) Briefly describe how you would set this up as a *supervised learning* task.

(b) Briefly describe how you would set this up as an *unsupervised learning* task.

(c) Briefly describe how you would set this up as a task for *SARSA-based Q-learning.*

(d) For each of the above, discuss one (1) relative strength and one (1) relative weakness of your approach compared to the others (you should compare to the others as a *group*; i.e., list $3 \times 2 = 6$ total advantages and disadvantages).

B766 – COMPUTER VISION BASIC QUESTIONS

B766-1. A grayscale image, $I(x,y)$, is to be smoothed by convolution with a discrete approximation of a 2D Gaussian kernel, $G_\sigma(x,y)$, of size (2k+1) × (2k+1) pixels.

(a) Give an expression for computing the intensity of a smoothed pixel, $S(x,y)$.

(b) Show how the convolution can be performed by two discrete 1D convolutions, and comment on the computation time savings this gives compared to using one 2D convolution.

(c) Describe and compare the *Laplacian-of-Gaussian* and *Canny* operators for detecting and localizing intensity edges in grayscale images.

(d) Give an expression for computing the directional derivative of I in the direction **n**, where **n** is a unit vector.

B766-2.

(a) Given an image in which a set of point features has been detected, describe an algorithm that uses RANSAC to detect k lines in the image, where k is given and most of the point features occur on or near at least one of these lines, though there are also many spurious (i.e., noise) feature points that do not occur on any of the k lines.

(b) Briefly describe what would be required to extend your method in (a) to determine the *best* number of lines (i.e., k is not given) in an image.

B776 – BIOINFORMATICS BASIC QUESTIONS

B776-1. When modeling a class of sequences, it is often important to model the distribution over the *lengths* of the sequences, in addition to representing the *composition* of the sequences. Suppose we are modeling DNA sequences and using Markov models with emission parameters.

(a) Draw the topology of a model for which the total probability over all sequences of length n is assigned as follows: $\Pr(length = n) = \theta \times (1-\theta)^{n-1}$, where $n > 0$ and $0 < \theta < 1$. Show the states, the transitions, and the transition parameters for the model. You do not need to show emission parameters, but indicate which states, if any, are silent.

(b) Draw the topology of a model for which the total probability over all sequences of length n is assigned as follows:

$$\Pr(length = n) = \begin{cases} 0, & \text{if } n = 1 \\ 0.6, & \text{if } n = 2 \\ 0.4 \times \theta \times (1-\theta)^{n-3}, & \text{if } n > 2 \end{cases}$$

Again, show the states, the transitions, and the transition parameters for the model. You do not need to show emission parameters, but indicate which states, if any, are silent.

B776-2. One limitation of standard clustering methods is that they are susceptible to local optima because they employ greedy search methods. One general strategy for finding better optima is to run a search process multiple times ensuring that a different search trajectory is explored on each run. Describe how you might use this strategy with <u>both</u> *hierarchical clustering* and *EM clustering*. In particular, answer the following questions about each modified algorithm.

(a) How does the method ensure that the search trajectory is different on each run?

(b) How does the method select the best clustering across the various runs?

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. Ensemble methods such as *Bagging* are popular because they often provide more accurate models than the method used to learn the individual members of the ensemble. Another important idea in machine learning is the heuristic of *Occam's Razor*.

(a) Describe the idea of Occam's Razor and explain why it's a useful idea.

(b) Is Occam's Razor a provably optimal criterion?

(c) Discuss why the success of Bagging might appear to be at odds with Occam's Razor.

(d) Discuss how we might reconcile these two ideas. That is, how can we coherently explain why and when both Bagging and Occam's Razor are useful in practice?

A760-2. The standard sources of "teacher" feedback in machine learning are labeled examples and reinforcements. However, for many situations, collecting such information can be challenging or even impossible. Fortunately, many other sources of training information frequently exist, and alternate problem formulations have been devised.

(a) Briefly describe the task of *multiple-instance learning* and explain how it broadens the applicability of machine learning.

(b) Assume that for a given domain a human teacher can provide some partial and possibly incorrect "domain knowledge" represented as Horn clauses. Explain how this can improve the effectiveness of machine learning.

(c) How does having the user provide Horn clauses address the *bias-variance dilemma*?

(d) Assume that you have to solve a reinforcement-learning task in robotics. Also assume that your system is instrumented in such a manner that you can collect "state-action" pairs while a human manually operates the robot 9am − 5pm each day. Describe how you might perform "learning from observation" in this scenario (it is fine if your solution mixes in some standard reinforcement learning as well, say outside of the 9am − 5pm period).

A766 – COMPUTER VISION ADVANCED QUESTIONS

A766-1. Consider the problem of *tracking* the 2D position and 2D orientation of a robot on the floor of a factory. A 2D map of the floor is given that specifies if a given point on the floor is empty space or is occupied by some object. A 2D image of the textured ceiling is also given, specifying the brightness at each point. As the robot moves, approximate motion information is available in the form of a vector $\mathbf{u}_t$ that specifies the motion from time t-1 to time t. Also, at each time t a camera on the robot that is pointed straight up at the ceiling computes the average brightness in a 10 x 10 window in the center of the image, specifying an observed brightness value, z_t. Given this information, describe how the CONDENSATION algorithm could be used to track the position and orientation of the robot as it moves around the floor. In particular, describe

(a) the components of an appropriate state vector, **x**, and the size of the complete state space assuming there are 100 possible values for each element of the state vector.

(b) a way to specify the initial sample-set of state vectors, $\{\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \ldots, \mathbf{s}^{(N)}\}$, assuming the initial 2D position and 2D orientation of the robot is completely unknown.

(c) the main steps in one iteration of the algorithm, applied to this particular problem.

(d) one of the main advantages of the CONDENSATION algorithm over other tracking methods.

A766-2. Consider the problem of *camera calibration* when the camera is to be used as a video surveillance camera that computes the speed of cars as they pass by a bridge where the camera is mounted. Assume the road is flat and straight and runs perpendicular to the bridge. The camera is mounted on the bridge in line with and parallel to the lane in which cars will be monitored. The lane has painted markings on it every 10 meters, exactly.

(a) What is the relationship between the pixel coordinates (u, v) corresponding to known 3D scene coordinates (X, Y, Z) for this problem where the known locations of the markings mean that for any fixed value of Y = c, there are markings at scene coordinates (0, c, 0), (10, c, 0), (20, c, 0), …, (50, c, 0)? Clearly describe the camera model you use.

(b) What is the minimum number of calibration points needed to solve this problem?

(c) Describe briefly how a linear solution to this problem would be set up.