FALL 2001
COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN – MADISON
PH.D. QUALIFYING EXAMINATION

Artificial Intelligence

Monday, September 24, 2001
3:00 - 7:00 PM
2317 Engineering

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:

- either one (1) of the questions G540-1 or G540-2, *and*

- 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 *seven* (7) 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 either one (1) of the questions G540-1 or G540-2

G540 – Introduction to AI Questions

G540-1. Consider the following version of the Towers of Hanoi problem: Given two disks of different diameters on a peg, *P1*, with the smaller disk, *D1*, on top of the larger disk, *D2*. There are two other pegs, *P2* and *P3*, that initially have no disks on them. The goal is to move the disks one at a time from peg *P1* to peg *P3* making use of all 3 pegs but never stacking the larger disk on top of the smaller one.

(a) Draw the state space graph for this problem. Label each state as a pair (*Pi*, *Pj*) indicating the peg that disks *D1* and *D2* are currently on, respectively. Also, indicate the start and goal nodes.

(b) If instead of 2 disks there are *n* disks and 3 pegs, which of the following heuristics are admissible for this problem. For each, either explain why it is admissible or give a counterexample if it is not. Disks are named from smallest to largest so that *D1* is the smallest disk and *Dn* is the largest disk. Hint: The optimal number of moves from the start state is $2^n - 1$.

i. The total number of disks on pegs other than the goal peg, *P3*.
ii. $2^k - 1$ where *Dk* is the largest misplaced disk.

(c) Using the heuristic function defined in (b)(i) and assuming all moves have unit cost, show the sequence of the first *five* (5) nodes expanded by best-first search using the evaluation function $f(n) = g(n) + h(n)$. If there are ties, specify how you chose to break ties.

G540-2. Express the following English sentences using situation calculus.

(a) Wherever a horse goes, her rider also goes.

(b) Selling a book does not change who wrote it.

(c) In the fall, trees lose their leaves.

(d) Unless they are glued down, blocks can be moved from one table to another.

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 – Basic Questions

B731-1. Consider a blocks world that consists of three blocks: *a*, *b*, and *c*. Suppose we are in an initial state in which *a* and *c* are on the table, and *b* is on top of *a*.

(a) What other facts must be true in this initial state?

(b) Draw two steps (3 "state" levels and 2 "action" levels) of the GRAPHPLAN planning graph from this initial state.

(c) Based on your planning graph in part (b), can we achieve *ontop(a,c)* from this initial state in two steps (i.e., with a sequence of two actions)?

The following is a review of the blocks world in case it is needed.

A state of the world is described by the following predicates.

Predicate	**Meaning**
ontop(X,Y)	*X* is on top of *Y*
ontable(X)	*X* is on the table
clear(X)	nothing is on top of *X*
holding(X)	the robot hand is holding *X*
handempty()	the robot hand is empty

The operators in the blocks world are the following, listed with their preconditions and effects.

Operator	**Preconditions**	**Effects**
pickup(X)	*clear(X), ontable(X) handempty()*	*holding(X), ~ontable(X) ~handempty()*
putdown(X)	*holding(X)*	*ontable(X), ~holding(X), handempty()*
unstack(X,Y)	*clear(X), ontop(X,Y), handempty()*	*holding(X), ~ontop(X,Y), clear(Y), ~handempty()*
stack(X,Y)	*holding(X), clear(Y)*	*ontop(X,Y), ~clear(Y), handempty(), ~holding(X)*

B731-2. Consider the task of learning a Bayesian network from data, where each data point is a complete setting of all the variables (there are no hidden variables and no missing values).

(a) Describe an algorithm for learning the structure of the Bayesian network from data. Be sure to make clear in your description the details of the state space, evaluation function and termination conditions that the algorithm will use.

(b) Discuss ONE major strength and ONE major weakness of your algorithm.

B760 – Machine Learning Basic Questions

B760-1. Overfitting avoidance is a central issue in supervised machine learning.

(a) Define *overfitting*.

(b) Explain its significance.

(c) For each of the following learning methods, describe and motivate a commonly used technique for overfitting avoidance:

i. Decision-tree induction
ii. Neural network training
iii. Naïve Bayesian learning
iv. Instance-based learning

B760-2.

(a) Define and informally motivate the basic PAC learning model.

(b) Informally explain the *VC dimension* and its significance.

(c) What is the VC dimension of the concept class *C* consisting of all disjoint pairs of intervals on the real line? Explain your answer.

$$C = \{a < x < b \ \acute{U} \ c < x < d \mid a, b, c, d \ \hat{I} \ \hat{A} \ \text{and} \ b < c\}$$

(d) Discuss the differences between the *mistake bound model* and the PAC model.

B766 – Computer Vision Basic Questions

B766-1. Describe the steps of an algorithm that uses the Hough Transform to detect dark rectangles of arbitrary size and orientation, on a lighter background in an image. Include in your answer all steps that precede and follow the steps associated with the Hough transform. Use a single parameter space, and give a description of its properties. Assume the use of an edge operator that gives for each pixel an edge magnitude and orientation.

B766-2. Consider the problem of 2D model-based object recognition where the goal is to find if there exists at least one occurrence of a given 2D object in an image. The object can occur in any 2D position, orientation and scale. The 2D object model is given as a set of *n* point features where each point's coordinates are given relative to the centroid of these points. Assume we have detected *m* point features in the input image, and assume that any image point feature could correspond to any model feature.

(a) Briefly explain why use of the RANSAC robust estimation method is a good approach for this problem.

(b) What is the minimum number of candidate corresponding points needed from the model and the image in order to initially estimate the 2D pose of an object in the image?

(c) How is the "goodness" of a hypothesized pose (as determined by a minimal set of point correspondences) computed by RANSAC for this problem?

(d) How is the best match pose determined by RANSAC for this problem?

B776 – Bioinformatics Basic Questions

B776-1. Consider using a hidden Markov model (HMM) for the following problem. We are given strings composed from an alphabet of two characters, *A* and *C*. We know that these strings are composed of alternating segments of two different types, *type 1* and *type 2*. We don't know the defining characteristics of these two types of segments, but:

- we have a set of training sequences that have been labeled with the corresponding segments in each,
- we know that *type 1* segments always begin with *AA* and *type 2* segments always begin with *CC* (however *AA* words can occur in *type 2* segments and *CC* words can occur in *type 1* segments).

(a) Draw a picture showing the graphical structure of a second-order Markov model that we could use to segment previously unseen sequences. Show all of the states and transitions in your model. Hint: use a model that has only transition probabilities.

(b) Briefly describe why this is considered a *hidden* Markov model.

B776-2. Consider a setting in which we will encounter a set of clustering problems. For each problem, we will be given a data set in which the instances are described by Boolean features. In some data sets, however, some of the instances will also have class labels associated with them. In other words, some of the clustering problems will involve only unlabeled data and some will involve a mix of labeled and unlabeled data. We can think of the labels as representing the "true" clusters for those instances that have them. Describe how you would adapt an existing clustering algorithm so that it was well suited to this setting.

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 *bias-variance* trade-off in the context of supervised learning.

(a) Define the bias-variance trade-off and briefly state its significance.

(b) In the case of learning with the *k*-nearest neighbor method, describe one way in which the trade-off between bias and variance can be controlled (i.e. the learner can be pushed toward either high-bias or high-variance models).

(c) In the case of learning with neural networks, describe one way in which the trade-off can be controlled.

A760-2. You work for a large software distribution company. The company gives out 5000 free copies of each new program as their chief method of advertising. (The company goal is to advertise by word of mouth of satisfied customers.) Your job is to decide which of the 1 million customers in your database should be given free software. Your database lists every software purchase or gift, including the date, for the last 10 years. The database also records which other customer(s) first told this customer about the product he/she purchased. These latter records are based on customer responses and hence are incomplete. Based on this data, you would like to predict who is influencing whom to buy software. If you can do this accurately, you can give free software to the most influential customers.

(a) How can you use machine learning to predict who is influencing whom to buy software? Be as specific as possible about your algorithm(s) and approach.

(b) What would you do differently if you also could "experiment" for a year before your boss starts to evaluate your performance? Assume that a year would permit you to test-market 100 software products.

Notes: For simplicity assume all the software products you market are of the highest quality, so that "negative" recommendations are not a problem. Also for simplicity assume that all products have equal cost, and every customer is potentially interested in every product. If you need to make other assumptions, please state them explicitly.