FALL 2002
COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN – MADISON
PH.D. QUALIFYING EXAMINATION

Artificial Intelligence

Monday, September 23, 2002
3:00 - 7:00 PM
Room 382, Mechanical 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:

- 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

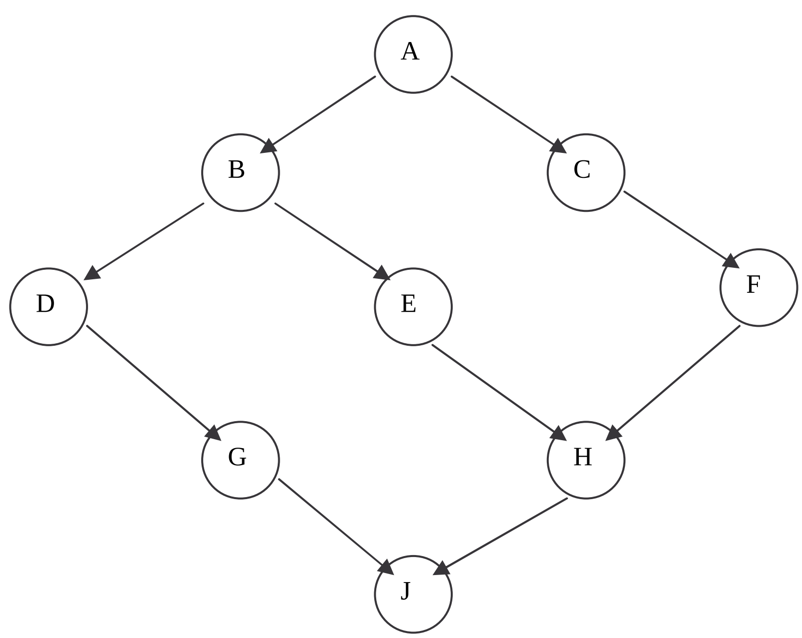
B731-1. A heap of integers is a binary tree in which each node contains an integer, and every integer is less than both its children.

(a) How can you represent a heap as a Prolog list?

(b) Write a Prolog program that takes as input an arbitrary list of integers and produces as output a heap containing exactly those integers.

(c) The algorithm heapsort takes as input an arbitrary list of integers, constructs a heap, and extracts the top (smallest) integer. It then re-builds the heap from the remaining integers, and recursively repeats until all integers have been extracted in order. Write a Prolog implementation of heapsort. Feel free to make use of any code you wrote for your answer to part (b).

B731-2. Given the following Bayes Net, construct the corresponding junction tree (ignore conditional probability tables).



B760 – MACHINE LEARNING BASIC QUESTIONS

B760-1. Using information theory is a common approach for choosing the best feature to next add to a decision tree being induced in a "top-down" manner.

(a) Show the calculations used for this task by Quinlan and informally explain their motivation. You may assume a two-class task in this part of the question.

(b) Now formally derive Quinlan's formulas using the principles of information theory. Here you need to handle the general case of K possible output classes.

(c) Explain one (1) way one might perform "bottom-up" induction of decision trees and informally discuss whether or not Quinlan's use of information theory is applicable in this approach.

B760-2. Both Boosting and rule-induction "covering" algorithms generate a series of "concept descriptions" when learning on a single training set.

a) Briefly sketch how each approach operates.

b) What do you feel is the most important *commonality* between these two approaches? Be sure to justify your answer.

c) What do you feel is the most important *difference* between these two approaches? Again, justify your answer.

d) For each of these approaches, explain one (1) technique that could be used to reduce overfitting.

B766 – COMPUTER VISION BASIC QUESTIONS

B766-1. Consider a database of satellite images, where each image contains regions from one or more of the classes: water, forest, and wheat fields. The goal is to segment each image into regions that are connected and homogeneous in terms of their class. One possible method is the *n-cut segmentation algorithm*.

(a) Formulate this problem for use with the n-cut segmentation algorithm, specifying what corresponds to nodes and edges in the graph, an appropriate affinity measure, how the matrices used by the algorithm are defined, and any parameters that must be defined by the user.

(b) Describe the main steps of the n-cut algorithm, as applied to this particular segmentation problem, where the final result is a set of regions of the three land-use types.

B766-2. Consider a pair of pinhole cameras with parallel optical axes and the optical centers are at coordinates (-b, 0, 0) for the left camera and at (b, 0, 0) for the right camera. The focal length of both cameras is 1.

(a) Give two reasons why corresponding image points (i.e., the projections of a scene point (x, y, z) in to the two cameras) might have different measured brightness.

(b) Say a planar square is positioned in the scene so that its vertical axis is parallel to the vertical axis in the camera, but the square's horizontal axis is rotated by angle θ around the vertical axis from a fronto-parallel orientation. The center of the square is at coordinates (-b, 0, z_0) and its width is w. What is the projected width of the square along the x-axis in the left and right images?

(c) Does stereo work when the cameras are parallel and projection is orthographic? Briefly explain why or why not.

(d) Discuss the advantages and disadvantages of a parallel camera configuration such as the one described above versus a configuration where the two optical axes are coplanar and intersect at a scene point (x_v, y_v, z_v), i.e., both image planes are rotated by the same angle, θ, towards a point in the scene.

B776 – BIOINFORMATICS BASIC QUESTIONS

B776-1. Describe an approach for clustering a set of sequences into a predefined number, *k*, of clusters. You should assume that the sequences are composed from a fixed alphabet of characters (like DNA or protein sequences), and that they are of varying lengths. Be sure to describe how

(a) clusters are represented,
(b) clusters are initialized and then refined by your algorithm,
(c) a "test" sequence would be assigned to a cluster.

B776-2. Consider using a hidden Markov model (HMM) for the following problem. We are given strings composed from an alphabet of four characters, *A*, *C*, *G* and *T*. We know that these strings are composed of alternating segments of two different types, *type 1* and *type 2*. We do not 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 both *type 1* segments and *type 2* segments are always between three (3) and five (5) characters long.

(a) Draw a picture showing the graphical structure of a hidden Markov model that we could use to segment previously unseen sequences. Show all of the states and transitions in your model. Indicate which states have emission parameters and which states (if any) are silent.
(b) Briefly describe why this is considered a *hidden* Markov model.

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.

(a) What is an ROC curve (define the axes)?

(b) Name one advantage of ROC curves over accuracy estimates for comparing two learning algorithms.

(c) Suppose we have run two learning algorithms on the same task and generated ROC curves for the two methods. Neither ROC curve dominates the other, but the two curves have a single crossover point. Can you use this observation to construct an improved predictor? If so, how?

(d) Suppose we want to generate ROC curves for a covering-based, rule learner (e.g. FOIL). Describe two (2) approaches to generating ROC curves for this learner. Discuss one (1) advantage and one (1) disadvantage of each approach.

A760-2. Imagine that one is searching a huge space for a good model of a set of examples that are described using N discrete-valued features. (Be sure to read all the parts of this question before writing anything, since the algorithm you choose to write about in Part *(a)* will be used in all parts of the question.)

(a) Many algorithms using *hill climbing* to perform this search. State one (1) algorithm that does so and briefly explain its operation in terms of search (e.g., state representation, next-state generator, heuristic function, and the criteria for termination).

(b) How would you extend the algorithm you discussed in Part (*a)* to do *beam search*?

(c) Describe one way to intelligently search the space in a (possibly partially) *randomized* manner. (You must describe an approach other than the one expressed in Part *d*.) Explain one (1) weakness and one (1) strength of this approach compared to the approach you described in Part *b*.

(d) Now imagine that you repeatedly (i.e., K times) choose a random starting point (which can, but need not, be a starting state in your search space) and then perform hill climbing. Describe how one could use the experiences obtained in trials 1 through $i - 1$ to hopefully improve performance in trial i. Be sure to explain what sense(s) of "performance" your approach aims to improve.

A766 – COMPUTER VISION ADVANCED QUESTIONS

A766-1. Let $p = (u, v)$ and $p' = (u', v')$ be corresponding points in two images taken at the same time from two different cameras of a static scene.

(a) Describe how these two points are related by the Fundamental Matrix, **F**.

(b) What is the size of **F**, the number of degrees of freedom in **F**, the minimum number of point correspondences needed to estimate **F**, and give an intuitive explanation of what **F** relates between the two cameras.

(c) Explain how **F** can be used to compute the epipoles in the two views.

(d) Describe an algorithm for computing **F** from a set of point correspondences.

(e) What is a major source of error in your algorithm, and what types of modifications could be done to reduce this error?

A766-2. One method for motion segmentation was developed by Wang and Adelson for partitioning the images into a set of "layers."

(a) Describe the main steps of the Wang and Adelson algorithm for motion segmentation, focusing especially on their method for grouping similar points.

(b) What is the aperture problem and does it affect the performance of this algorithm?

(c) Describe two main weaknesses of this algorithm.

(d) Describe how this algorithm could be used or extended for reducing the camera shake from a hand-held video camera.