SPRING 2005
COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN – MADISON
PH.D. QUALIFYING EXAMINATION

Artificial Intelligence

Monday, January 31, 2005
3:00 – 7:00 p.m.
Room 2534 Engineering Hall

GENERAL INSTRUCTIONS:

a) Answer each question in a separate book.

b) 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.*

c) 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. Consider constructing a Bayesian network of four variables: *A, B, C, D.*

a) Draw a network over these variables such that, with no evidence at *B*, the variable *A* is *d*-separated from *C* but not *D*; with evidence at *B*, *A* is *d*-separated from *D* but not *C*.

b) In your Bayesian network for part (a), suppose the only evidence is at node *D*. Is *A* still *d*-separated from *C*?

c) Draw a junction tree for your Bayesian network. Label the nodes with clique names and the edges with separator names?

B731-2.   Suppose you wish to implement Quinlan's FOIL algorithm directly in a relational database management system, so that it requires as little input from the user as possible.

a) Describe the inner loop of the FOIL algorithm using pseudocode. Include brief comments.

b) To run FOIL on any existing database with multiple tables, in addition to the database what would your system require the user to specify?

c) What items required by the algorithm, as defined in (a), can your system get (or compute) directly from the database rather than requiring from the user as input.

B760 – MACHINE LEARNING BASIC QUESTIONS

B760-1. A key concern in supervised learning is *overfitting avoidance*.

a) Define *overfitting* and explain its importance.

b) Discuss <u>one</u> key technique (three in total) for addressing the problem of overfitting in <u>each</u> of the following approaches to supervised learning:

   i. Decision trees
   ii. Bayesian networks
   iii. Neural networks

c) Describe how optimization in support vector machines (for supervised learning) can be viewed as an application of the *minimal description length* principle. Be sure to include in your description a discussion of any relationships to overfitting.

B760-2. Imagine an environment for *reinforcement learning* that contains three light bulbs, *LB1*, *LB2*, and *LB3*. The legal actions are toggling the on-off switch on any one light bulb, with the exception that the task ends when all lights are on. The immediate-reward received is +1 if the number of "on" bulbs is even and -1 if the number of "on" bulbs is odd. Use a *discount rate* of 0.9 for all parts of this question.

a) You initialize a Q *table* to have Q=0 for all state-action pairs. You then start with all bulbs "off" and execute the policy "turn *on* the lowest numbered bulb that is off" until the task ends. Following this experience, you run the one-step (non-SARSA) Q-learning algorithm. Show what changes in the Q table.

b) Repeat Part (a), again starting with Q=0 for all state-action pairs, but this time use the two-step (non-SARSA) Q-learning algorithm.

c) This time you choose to use a *function approximator* to represent the Q function. For simplicity, you use a perceptron with a linear output unit. You initialize all free parameters to 0 in this perceptron. Which training examples would you construct from the experience in Part (a), using one-step Q learning? Show how training on the first of these examples would alter the perceptron. Use 0.1 as the perceptron's learning rate.

d) Describe one key advantage and one key disadvantage of using function approximators in reinforcement learning.

B766 – COMPUTER VISION BASIC QUESTIONS

B766-1. Curve Detection

a) Describe the main steps of a complete procedure for detecting a parabola in an image. The parabola will be composed of edge pixels, but may contain gaps at arbitrary points. The image may also contain edge points that are not part of the parabola. The equation for a parabola is $y = ax^2 + bx + c$.

   Your method should find the parabola that contains the most edge pixels in the image. Be sure and explicitly specify any formulas that are needed. Assume that the edge pixels have already been identified so that $I(x,y) = 1$ if (x,y) is an edge point, and 0 otherwise.

b) What is the running time of your algorithm? Define any symbols that you use.

c) Describe <u>one</u> significant advantage and <u>one</u> important disadvantage of your method. Possible topics to consider in your answer are ability to correctly detect parabolic curves, algorithm speed, and accuracy and robustness in determining the true position, orientation, and scale.

B766-2. Segmentation

Describe briefly the main steps in methods for finding planar regions in a disparity map (previously computed by some stereo algorithm) using (a) RANSAC, and (b) *k*-means clustering. Give one advantage of each compared to the other.

B776 – BIOINFORMATICS BASIC QUESTIONS

B776-1. There are well-established algorithms, such as Baum-Welch, for learning the parameters in hidden Markov models (HMMs). The problem of learning the **structure** (i.e., topology) of an HMM has not been investigated as much, however. Discuss how you might apply an AI-style search method to the task of learning HMM structure. In particular, describe:

a) the operator(s) used by your search algorithm,

b) the search strategy employed, and

c) the evaluation function you will use, if any.

B776-2. Describe how you might extend standard EM-based clustering to induce a two-level **hierarchical** clustering. That is, the clusterings returned by this method consist of a "lower" layer of clusters and an "upper" layer that characterizes relationships among the clusters in the lower layer. You can assume that each instance is an *n*-dimensional vector of real values, and the number of clusters at each layer is pre-specified.

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 *transferring* what was learned on some previous task, *Prev*, in order to do a better job of learning the current task, *Curr*. Assume both are supervised-learning tasks. The two tasks share some but not all features.

a) One possible way to do transfer is to somehow use some structure learned in *Task Prev* to improve learning for *Task Curr*. For any one of the approaches listed below, describe how you could use some of the learned structure for *Prev* when learning *Curr*. (Your answer should go beyond simply using the output of *Task Prev* as an input feature for *Task Curr*.)

   i. Bayesian networks
   ii. Decision trees
   iii. Neural networks
   iv. Support vector machines

b) Imagine you have the model learned for *Prev* and *k* training examples for *Curr*. Describe an experimental methodology that you could use to judge whether or not transferring something learned on *Prev* improved the learning of *Curr*. Informally sketch (on a standard "*y* as a function of *x*" 2D curve) some results that you expect to obtain from your proposed experiment for the case when transfer is helpful.

c) Now imagine you have *N* previously learned models - call them $Prev_1$ through $Prev_N$. Assume you want to choose only one of these to aid *Task Curr*. Describe and justify an algorithm for choosing which previous model to use.

A760-2. The most common way to represent conditional probability distributions (CPDs) in Bayesian networks is to use tables. Other representations may be used, however. Consider a situation in which decision trees are used to represent CPDs.

a) Assuming that you are given the set of candidate parents for each node in the network, briefly describe how you would apply a decision tree learning algorithm to learn the CPD for each node.

b) Now suppose you are trying to learn both the structure and parameters of the Bayesian network (i.e., you are not given the set of candidate parents for each node). Describe how you would extend your approach from Part (a) to this situation.

c) Discuss how you could trade off *bias* and *variance* in the models learned by your method. That is, how would steer your method toward a high-bias, low-variance solution or vice versa?

A766 – COMPUTER VISION ADVANCED QUESTIONS

A766-1. Consider adapting Ullman and Basri's object recognition method of using a linear combination of models to the problem of outlier detection in motion tracking. That is, given a sequence of image frames, assume corner points have been detected in each frame and then the points have been tracked from frame to frame resulting in a set of "trajectories." That is, each trajectory is a sequence of image coordinates, one pair of coordinates for each frame in the sequence, specifying the position of that trajectory in each frame of the sequence. Let (x_{fp}, y_{fp}) specify the image coordinates of the p^{th} trajectory in the f^{th} frame. The problem then is how to detect when a point in a trajectory is wrong in that it does not correspond to the same physical scene point as the other points in the trajectory, and therefore it is an "outlier" that should be removed.

a) What camera model is appropriate to assume for this problem? Explain why. How many parameters have to be solved for?

b) Give expressions for how the coordinates of a point in the i^{th} frame, (x_{ip}, y_{ip}), can be defined in terms of the coordinates of other points in the same trajectory, p, but at different "reference" frames.

c) Describe the main steps in a method that uses the Ullman and Basri linear combination idea to detect when a point in a given trajectory is an outlier.

A766-2. Consider the problem of detecting bronchiectisus, which are ring-like structures, in medical images of human lungs. The structure of these 2D objects is that there is a light ring surrounding a dark circular region. The width and diameter of the ring can vary over a known range, the inner dark region can vary in size, and the shapes of the ring and inner circle can vary as well so that they are more elliptical, though not perfectly so. In a single image there may be from 0 to 10s of occurrences of this type of "object." Describe some of the key issues that must be addressed in order to solve this problem. Then describe a method for solving this problem in a robust way (i.e., with few false positives and few false negatives detected).

This page intentionally blank. You may use it for scratch paper. Please note that this page will NOT be considered during grading.

This page intentionally blank. You may use it for scratch paper. Please note that this page will NOT be considered during grading.