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

Artificial Intelligence

Monday, February 4, 2002
3:00 - 7:00 PM
2311 Chemistry

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 problem of constructing crossword puzzles. That is, fitting words into a grid of intersecting horizontal and vertical squares. Assume that a list of possible words (i.e., a dictionary) is provided, and the goal is to fill in all the squares using any subset of the words in the dictionary.

(a) Describe all of the components of a state-space formulation of this problem including state representation, start state, goal, operators, etc.

(b) Define a heuristic function for this problem and explain what it does. Discuss whether or not your heuristic is admissible.

(c) Would a bi-directional A* search be a good idea for this problem? Briefly explain why or why not.

G540-2. Consider the following pairs of English sentences and proposed translations into first-order predicate calculus (FOPC).

(a) *John sees a woman with a telescope.*
$\exists X \ \text{woman}(X) \wedge \text{sees}(\text{john}, X) \wedge \text{possesses}(X, \text{telescope})$

(b) *Some books are long.*
$\forall X \ \text{book}(X) \rightarrow (\text{long}(X) \vee \neg \text{long}(X))$

(c) *Blue cars cost more to buy than red cars do.*
$\forall X, Y \ [\text{car}(X) \wedge \text{car}(Y) \wedge \text{blue}(X) \wedge \text{red}(Y)] \rightarrow \text{greater}(\text{price}(X), \text{price}(Y))$

For each pair, which of the following possibilities is the *best* evaluation of the translation? Explain your answers, using formal *interpretations* where appropriate.

(i) There is a serious flaw in the translation from English to FOPC.

(ii) The translation into FOPC does not include some of the ambiguity present in the English, but is otherwise correct.

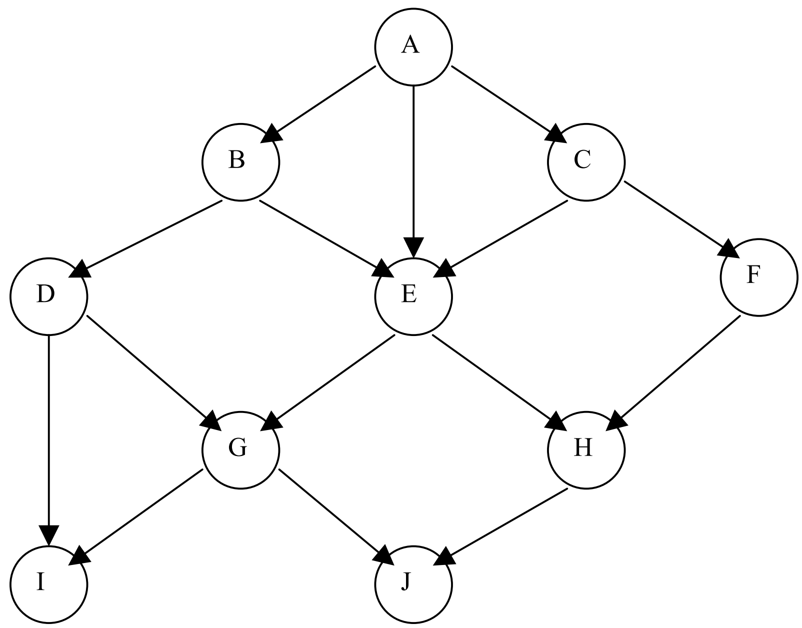
(iii) The FOPC translation is *more* ambiguous than the original English.

(iv) The translation from English into FOPC is correct.

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. Given the following Bayes Net, construct the corresponding junction tree (ignore conditional probability tables and just show the graph of the junction tree).



B731-2.

(a) Describe the GSAT algorithm.

(b) Choose two (2) other AI search strategies, and give two (2) significant ways in which GSAT differs from each of these.

(c) Give one (1) advantage and one (1) disadvantage of GSAT compared with each of these other two search strategies.

B760 – Machine Learning Basic Questions

B760-1. Consider the task of choosing a good set of features for supervised learning.

(a) One approach to feature selection is to use *information gain* to choose the top *N* features. Discuss one (1) strength and one (1) weakness of such an approach.

(b) Another idea proposed in the literature is to use *forward* or *backward* selection of features. Explain this approach and describe one (1) strength and (1) weakness of forward/backward selection as compared to the approach outlined in Part (a).

(c) A third idea is to first use a decision-tree algorithm to fit the training data and then use as one's feature set only those features that appear in the induced decision tree. How does this approach compare to Part (a)'s, especially given that both use information gain? Explain one (1) strength and one (1) weakness as compared to Part (a)'s approach.

(d) Assume that you employ the approach in Part (c) and then use the chosen features with the naïve Bayes algorithm. Choose one (1) major weakness of naïve Bayes and discuss whether or not the feature-selection step might overcome this weakness.

B760-2. Consider the approach of using Q learning to address the reinforcement-learning task.

(a) One appealing property of Q learning is that this method internally creates the "input-output" pairs needed by supervised learning. Explain.

(b) Explain why "rote memorization" of these input-output pairs is still an interesting form of machine learning.

(c) Describe the steps that need to be undertaken to use artificial neural networks to learn the Q function. What are the one (1) major strength and one (1) major weakness of using neural networks for Q learning?

(d) Explain what changes to the standard ID3 (or C4.5) decision-learning algorithm are necessary in order to use it to learn the Q function. (No need to explain more than two (2) major changes.)

B766 – Computer Vision Basic Questions

B766-1. Compare the Canny edge detector and the Laplacian-of-Gaussian (LoG) edge detector for each of the following questions:

(a) Which of these operators is/are isotropic and which is/are non-isotropic?

(b) Describe each operator in terms of the order of the derivative that it computes.

(c) What parameters must be defined by the user for each operator?

(d) Which detector is more likely to produce long, thin contours? Briefly explain.

(e) Which detector is better for detecting corners? Briefly explain.

B766-2. Consider the problem of 2D object recognition from a single image where the object may occur in any 2D position, 2D orientation, or scale in the image. Assume the image is captured by a pinhole camera under orthographic projection and there are m point features that define the object model, and n point features detected in the input image.

(a) How many dimensions are there in the Pose Space for this problem? Briefly describe what each dimension means.

(b) How many point correspondences are needed to completely determine the mapping between the object model and an instance of it in the image?

(c) Describe a Pose Space search technique that is analogous to the voting technique used by Thompson and Mundy (they used vertex-pair features for determining pose, but here we're using point features). How many votes will be cast in Pose Space? Briefly explain.

(d) If $n >> m$, why is this method likely to work poorly? Suggest a way to overcome this problem.

B776 – Bioinformatics Basic Questions

B776-1. Suppose you wish to train an HMM to detect when a user of a computer account is working on a scientific paper. Assume that the sequence of commands (without their arguments) issued by the user from a UNIX command line is recorded (e.g., *..., mv, xemacs, netscape, latex, ls*, ...). You are given training sequences consisting of long sequences of commands. Certain points in the sequences have been labeled as indicating (i) when the user began working on a paper, (ii) when the first draft was finished, and (iii) when the final draft was finished.

(a) Draw a diagram illustrating the topology of an HMM for this task. Describe what the emission and transition parameters in the model represent.

(b) Describe how you would use the HMM to process test sequences, predicting the points in the sequence when the user began working on a paper and when the paper was finished.

(c) How would you adapt your approach to be able to detect when the user was working on two papers at the same time? You can assume that you are given training sequences representing this case.

B776-2. Consider the problem of trying to select the "right" number of clusters when clustering a given data set. Describe an approach to doing this for some clustering algorithm. State which clustering algorithm your method is designed for, and how the number of clusters is controlled in this algorithm. Describe what your method is trying to optimize and how it is computed. Discuss what assumptions your method is based on.

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. Suppose we wish to train a system to classify images according to whether they contain houses or not. We are given a training set with images labeled as either positive or negative instances. A positive instance is one that depicts a street-level view of a house *somewhere* in the image.

(a) Describe how you would apply a *relational* learning method to this task. What are the objects you would use to represent each image? What are the relations among objects that your representation would include?

(b) Suppose that your training (but not test) instances were also labeled with information about how many doors and windows were visible in each positive instance. How would you adjust your training procedure and/or representation to take advantage of this information?

(c) Briefly describe one (1) significant strength and one (1) significant limitation of using a relational method for this problem.

A760-2. Suppose that you have a file containing the measurements of 1000 numeric features from each of 10 tumor and 10 normal-tissue samples. You apply boosted decision trees to build a classifier to distinguish between *normal* and *tumor*.

(a) What might indicate that your boosted-tree algorithm is *overfitting*? Specifically, give two (2) different indicators and explain why they are indicative of overfitting.

(b) Why are boosting algorithms prone to overfitting?

(c) Describe one (1) alternative ensemble approach that you could use to avoid overfitting. Explain why the alternative you selected might be less prone to overfitting.