SPRING 2000
COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN – MADISON
PH.D. QUALIFYING EXAMINATION

Artificial Intelligence

Tuesday, February 8, 2000
3:00 – 7:00 PM
226 Noland 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:

- either one (1) of the questions G540-1 or G540-2, *and*
- both (2) questions in the section labeled B760, B766, or B780 corresponding to your chosen focus area, *and*
- any two (2) additional question in the sections B760, B766, and B780, 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.

Suppose we wish to use *first-order predicate calculus* (FOPC) to build a reasoning system for diagnosing faulty computers. Write down and briefly explain four (4) FOPC rules that might be of use for such a system. At least two (2) of your rules must use situation calculus to represent diagnostic actions.

G540-2.

Lung cancer does not cause smoking. So why does discovering that someone has lung cancer cause us to suspect they may be a smoker?

Suppose we know that the probability of contracting lung cancer if one is a regular smoker is 0.8, and that we know the prior probability of someone being a smoker is 0.1. Assume that we do not know the prior probability of contracting lung cancer, but that the probability of contracting lung cancer for non-smokers is 0.001. What is the probability that someone is a smoker given that they have lung cancer?

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).

B760 – Machine Learning Basic Questions

B760-1.

Suppose we want to build a system that automatically categorizes and files the incoming email of a given user. This system will look at the header and content of an incoming message and then file that message into one of several mail folders: *meetings*, *junk-mail*, *personal-mail*, etc. Sketch the design of a system component that learns to classify mail messages according to the folder in which the user would most likely put the message. Training data for this task will consist of email messages that have been manually categorized by the user. Your learning system should take into account that (a) additional training data will be provided each day, and (b) the user's mail-organizing behavior might change over time (e.g. something previously considered interesting might now be considered junk mail). Be specific in describing the features that your system will use to represent messages, how your system will make use of a continual stream of training data, and how it will adapt to the user's behavior changing over time.

B760-2.

A decision stump is a decision tree with exactly one internal node. Given the following data set, what decision stump would be returned by an algorithm which chooses the split that maximizes accuracy on the training set?

Show how you could use three-fold cross-validation to estimate the future accuracy of this algorithm on this dataset.

x1	*x2*	*x3*	*x4*	*class*
1	0	1	0	+
1	1	1	1	+
0	1	1	0	+
0	0	1	1	+
1	1	0	0	+
1	0	0	1	-
0	1	0	1	-
0	1	1	1	-
0	1	0	0	-
1	1	0	0	-

B766 – Computer Vision Basic Questions

B766-1.

(a) Explain, giving any definitions appropriate:

(i) the epipolar geometry of a binocular camera system, including reference to the epipoles and epipolar lines

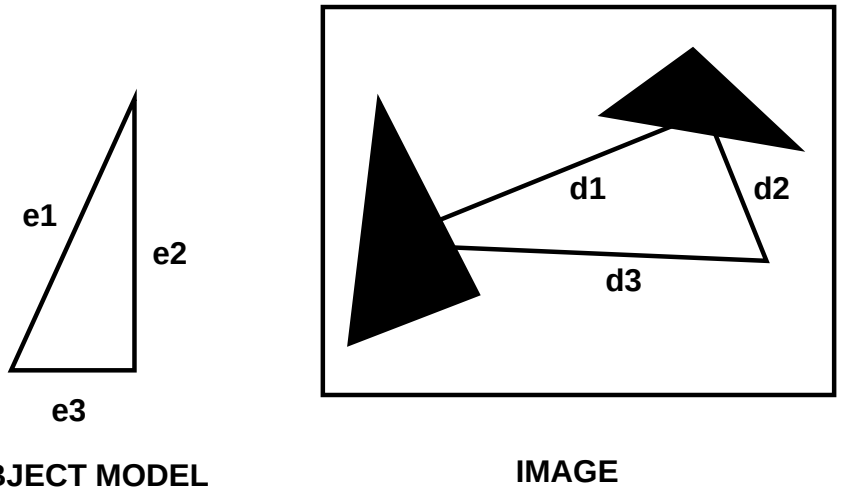
(ii) disparity

(b) Assume a parallel camera geometry with a baseline of 30cm. Each camera has a focal length of 20mm and each pixel is 0.1mm by 0.1mm square. Calculate the world coordinates of a point whose image coordinates (in pixels relative to the center of the image) are $(20, 10)$ in the left image and $(10, 10)$ in the right image (assume that 1mm is 10 pixels).

B766-2.

(a) Define and describe in detail "model-based recognition." Include in your description reference to the major components of a recognition system, defining any terms you introduce.

(b) Using the fact that angles between pairs of line segments in a frontal view (i.e., when the plane containing the line segments is perpendicular to the optical axis) are invariant to either orthographic or perspective projection, construct a complete interpretation tree for matching the image data (**d1, d2, d3**) to the object model (**e1, e2, e3**) shown in the figure below.



B780 – Robot Motion Planning Basic Questions

B780-1.

A point mobile robot knows its own and its goal coordinates at all times, and has a tactile sensor that allows it to sense obstacles it touches. Does the following sensor-based motion planning algorithm converge? Prove your response and make sure all possible cases are covered by your proof.

Step 1. Move toward the target along a straight line until one of the following occurs:

(a) The target is reached; the procedure stops.

(b) An obstacle is encountered and a hit point, H, is defined; go to step 2.

Step 2. Turn left and follow the obstacle boundary. If the target is reached, stop. Otherwise, after having traversed the whole boundary and having returned to H, define a new leave point L (the point on the obstacle boundary that is closest to the target point). Follow the boundary back to point L. Go to step 1.

B780-2.

Consider a two-dimensional, two link PR arm manipulator (P = "prismatic", R = "revolute"). Derive the inverse kinematics equations for this arm. How many solutions for the two joint angles are possible? What is the shape of this arm's work space? Draw a sketch or explain.

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.

Combining the outputs of multiple classifiers often results in more accurate predictions than those made by any of the individual classifiers. Of course, for this to happen, the individual classifiers must differ somewhat in the functions they represent. *Bagging* and *boosting* ensure diversity among the classifiers by giving each one a different training set, or a different distribution over the training set. Sketch an alternative approach to learning diverse classifiers that does not involve different training sets for the individual classifiers. Describe the assumptions made by your approach about the nature of the tasks and the learning algorithms to which it can be applied, as well as the number of individual classifiers it can produce.

A760-2.

It can be very expensive to collect a sufficient number of labelled examples for supervised learning. On the other hand, while large number of unlabelled examples are often available, unsupervised learning can be insufficiently guided.

(a) Propose and justify a technique for augmenting a small set of labelled examples with some unlabelled ones. State any assumptions you are making about the labelled and unlabelled examples.

(b) Briefly explain one (1) "real world" task where it might make sense to apply your approach.

(c) Describe the experimental methodology that you would use to evaluate your approach.

(d) Explain how the *theory refinement* approach addresses the high cost of obtaining labelled examples.