SPRING 1996
COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN – MADISON
Ph.D. QUALIFYING EXAMINATION

Artificial Intelligence

MONDAY, FEBRUARY 5
3:00 – 7:00 PM
ROOM 2104 CHAMBERLIN

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, B761, or B766 corresponding to your chosen focus area, *and*
- any two (2) additional question in the sections B760, B761, and B766, where these two questions need *not* come from the same section, *and*
- both (2) questions in the section labeled A76x 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. Explain briefly what best-first *search* is in general. Now, suppose a computer program is trying to use binary resolution to derive the empty clause from the clauses:

$\neg(x < y) \vee \neg(y < z) \vee x < z$
$\neg(x < x)$
$a < b$
$b < c$
$c < d$
$d < a$

Here, x, y, z are variables and a, b, c, d are constants. Explain how best-first search could help the program find the proof more quickly than would a naive breadth-first search.

G540-2. There are a number of courses in the CS Department at UW-Middleton, with many required for graduation. Below are all the courses you must take to graduate, not in any particular order. Starting with the course at the top and going to the bottom, add each course to a plan representation in the same way that a *partial-order (i.e., non-linear) planner* would. Draw the plan at each step as it evolves and say briefly why each step goes where you put it. Assume that you start having taken no courses.

OPERATOR	PRECONDITION	EFFECT
$cs16$	$cs15\_credit$	$cs16\_credit$
$cs17$	$cs32\_credit$	$cs17\_credit$
$cs32$	$cs16\_credit \wedge cs15\_credit$	$cs32\_credit$
$ee10$		$ee10\_credit$
$cs19$	$cs16\_credit$	$cs19\_credit \wedge \neg cs15\_credit$
$cs15$		$cs15\_credit$
$graduation$	$cs19\_credit \wedge cs17\_credit \wedge ee10\_credit$	$graduated$

Answer both (2) of the questions in the section labeled B76x 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. Sketch the proper methodology for using *cross validation* to both (a) set the parameters in an inductive learning algorithm and (b) estimate the future accuracy of the induced concept. Describe one (1) possible methodological flaw that a sloppy experimenter might make and explain why it is incorrect. Finally, briefly but concretely sketch one (1) possible use of cross validation for parameter setting in both (a) a genetic algorithm and (b) a neural network trained by backpropagation.

B760-2. Define the general task of *reinforcement learning*, then apply the *Q-learning* algorithm to the simple testbed specified below.

- The environment is a 2x2 grid of squares. The learner can move horizontally or vertically, but not diagonally.
- The learner always follows its *policy*, breaking ties randomly. (It initially estimates that the utility of each state is zero.)
- The learner starts each "game" in the lower-right cell.
- The upper-left square always contains some "cheese," and whenever the learner enters this cell a reward of 10 is received; all other moves produce a reward of -1. Capturing the cheese ends this episode, and the game restarts. Assume a discount rate of 0.9.

Show the initial *Q-table* for this scenario, then explain how this table evolves through each of the first three (3) steps of the learner. Finally, briefly describe one (1) strength and one (1) weakness of using tables in Q-learning.

B761 – Deductive Problem Solving & Logic Programming Basic Questions

B761-1. Convert the following logical sentences to clauses, and then derive the empty clause, using *binary resolution plus paramodulation.*

$\exists y \forall z (y < z \;\Rightarrow\; f(z) = g(z))$
$\exists y \forall z (y < z \;\Rightarrow\; g(z) = h(z))$
$\neg \exists y \forall z (y < z \;\Rightarrow\; f(z) = h(z))$
$\forall x \forall y \exists z (x < z \wedge y < z)$
$\forall x \forall y \forall z \; ((x < y \wedge y < z) \;\Rightarrow\; x < z)$
$\forall x (x = x)$

B761-2. Consider the *Prolog* program:

```
unemployed(X) :- not works_at(X, Job).
works_at(bill, presidency).
works_at(bob, senate).
```

The query `?- unemployed(bill)` will result in a `no` in most Prolog's (in Quintus Prolog, you would use a "\+" instead of a "`not`" in the program). Explain this behavior in terms of the *completion* of the program.

B766 – Computer Vision Basic Questions

B766-1.

(a) In what types of computer vision applications would *template matching* be an appropriate method for two-dimensional object recognition? Give your answer in terms of general domain characteristics and problem degrees of freedom.

(b) What are two significant limitations of template matching?

(c) One approach to template matching uses *cross-correlation* as a measure of match. Compute the cross-correlation between the following grayscale image f and template t. Assume both have 0 values outside of the 3x3 windows that are shown.

$$\begin{matrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{matrix} \qquad\qquad \begin{matrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{matrix}$$
$$f \qquad\qquad\qquad\quad t$$

(d) Give an example of another image g that intuitively should match worse than the image f in (c), but in fact the cross-correlation of g with t at at least one point is better than the best matching point when cross-correlating f with t.

(e) How can the problem in (d) be overcome?

B766-2.

(a) Describe the four main steps in the *Canny edge detector*. Also, indicate the purpose of each of these steps.

(b) Why might one prefer to use the Canny operator instead of the *Laplacian-of-Gaussian* edge operator?

Answer both (2) of the questions in the section labeled A76x that corresponds to your chosen focus area.

A760 – Machine Learning Advanced Questions

A760-1. The most common approach to machine learning is for the teacher to simply provide a set of labelled training examples to an inductive learning algorithm. However, there are many other types of information that a teacher could provide. For each of the following approaches to machine learning, describe how one (1) additional form of training information could be used, referring to published work where appropriate; give one (1) concrete example of each additional type of training information that you describe.

- *inductive-logic programming*
- *nearest-neighbor algorithms*
- *reinforcement learning*
  (in this case, describe an addition to the traditional reinforcement signal)

Also, state a form of teacher-learner communication that is used when both the teacher and learner are humans, but has not yet been captured in a machine learning system. Sketch a plausible approach for creating a machine learner that does make use of this source of training information.

A760-2. Explain the *bias-variance dilemma*. Describe one (1) approach to learning-from-examples that has high bias and one (1) that has high variance. Do you see any significant relationship between this dilemma and the *minimal description length* (MDL) principle? Finally, discuss how (a) *ensemble* methods and (b) *domain-specific knowledge* can help address the bias-variance dilemma. (An *ensemble* is a set of trained classifiers whose outputs are somehow combined to produce a composite classification.)