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

Artificial Intelligence

Monday, September 19, 2005
3:00 – 7:00 p.m.
Room 1240 Computer Sciences Building

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 the task of learning Bayesian network conditional probability tables (CPTs) from data that may have missing values, including perhaps hidden variables.

(a) Describe EM and Gibbs sampling algorithms for learning Bayesian network CPTs given data with missing values and/or hidden variables.

(b) Describe one weakness of the EM approach. Discuss whether there exists a way to address this weakness within EM.

(c) Describe one weakness of the Gibbs sampling approach. Discuss whether there exists a way to address this weakness within Gibbs sampling.

B731-2. Suppose we wish to use FOIL to learn rules to predict whether a web page is a student web page.

(a) How would you represent information about web pages for FOIL? Specify all your predicates, including their arguments. Be sure to include your target predicate.

(b) Describe one advantage of a relational learner such as FOIL over a feature-vector learning algorithm for this task.

(c) Describe one advantage of FOIL over another ILP algorithm (your choice) for this task.

(d) Describe one disadvantage of FOIL compared with another ILP algorithm (your choice) for this task.

B760 – MACHINE LEARNING BASIC QUESTIONS

B760-1. Consider the use of ensembles in machine learning.

(a) Explain how ensembles address the issue of overfitting avoidance.

(b) What is one strength of bagging compared to boosting? Justify your answer.

(c) Give a brief argument for the use of an ensemble consisting of one decision tree, one support-vector machine (SVM), and one neural network *instead* of using an ensemble of three models all produced by the same learning algorithm.

(d) Imagine you plan to perform a 10-fold, cross-validation experiment involving bagging. Describe an experimental methodology for choosing the size of the ensembles you would use.

B760-2. Consider the task of reinforcement learning (RL).

(a) Describe a simple RL environment where suboptimal behavior results from always *exploiting* the current policy. Be sure to explain your answer.

(b) Imagine an RL environment with primitive actions *a, b, c,* and, *d,* each of which can be legally executed in any state in this environment. Assume you wish to learn and deploy useful "macro-operators," that is, compositions of the primitive actions. One such operator might be: *execute action* ***a*** *followed by* ***c*** *followed by* ***b***. Use *acb* to denote this macro-operator.

i. Assume you have learned a good set of macro-operators and wish to use them in combination with the primitive actions. Explain how you would use this expanded set of actions with the SARSA algorithm.

ii. Propose and justify an algorithm for learning a useful set of macro-operators.

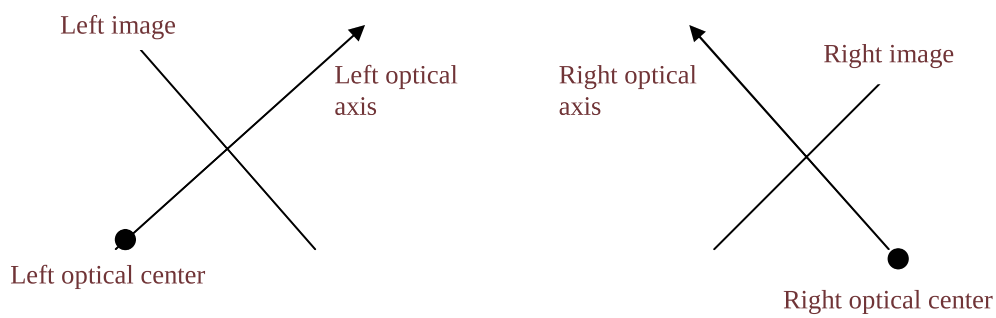
iii. What would be one major advantage of learning macro-operators?

iv. What would be one major disadvantage of learning macro-operators?

B766 – COMPUTER VISION BASIC QUESTIONS

B766-1.

(a) A stereo rig consists of a pair of cameras oriented so that their optical axes form an angle of 90 degrees as shown below; the *y* axes of the two cameras are parallel to each other and point out of the page. Draw the two images and sketch on them the epipolar geometry, indicating clearly the positions of the *epipoles* and the *epipolar lines*. Also, define these terms in general.



(b) If the left camera is rotated counter-clockwise about its optical center, with the axis of rotation the camera's *y*-axis, until the two optical axes are parallel, how does the family of epipolar lines in the right image change?

(c) Describe one advantage and one disadvantage of having the optical axes oriented as in (a) as opposed to having them parallel to one another.

(d) Define the *ordering constraint* (also known as the monotonicity constraint) that is used in stereo correspondence algorithms to disambiguate point matches on corresponding epipolar lines. Sketch a configuration of scene surfaces and cameras where the ordering constraint is valid, and a configuration where it is not valid.

B766-2.

(a) Define the main steps of the Normalized-Cut image segmentation method, including the quantity that is minimized, assuming it is being used to segment an image into exactly two regions.

(b) Why is the quantity being minimized *normalized* and what is a potential problem if it is *not* normalized?

(c) Say the image we are trying to segment corresponds to a graph with two connected components. Describe the meaning of the eigenvector corresponding to the largest eigenvalue.

(d) If the image contains more than two regions, two possible algorithms for segmenting it are (1) to use the algorithm in (a) recursively, and (2) to use multiple eigenvalues and eigenvectors.

i. Briefly describe how the second approach would work.
ii. When are these two algorithms likely to *not* give the same results? Justify your answer.

B776 – BIOINFORMATICS BASIC QUESTIONS

B776-1. Two key limitations of EM-based clustering are:

i. The number of clusters has to be specified in advance.
ii. The EM approach may find only a *local* minimum in the data likelihood.

For both of these limitations answer the following questions.

(a) Describe an approach that can, in part, address the limitation.

(b) Discuss one significant strength of each of your approaches.

(c) Discuss one significant weakness of each of your approaches.

B776-2. Consider the task of training an HMM or SCFG system to automatically extract information from sentences in biomedical articles. In particular, suppose we want our system to return for each sentence (i) the names of proteins referenced, and (ii) whether the sentence is *speculative* or not. For example, given the following sentence:

Sp1 may serve at least in part as a carrier to bring c-Jun to the promoter.

the system would identify *Sp1* and *c-Jun* as proteins, and classify the sentence as being *speculative.* To train the system, you are given a set of such sentences in which protein names are labeled in the text and the sentence as a whole is given a *speculative* or *not-speculative* label.

(a) Describe how you would train either HMMs or SCFGs for this task. Outline the structure of your model(s) and describe what learning would involve.

(b) Given your trained models, how would they be applied to extract information from reviews? What algorithms would you apply?

(c) Briefly discuss one significant advantage of using HMMs instead of SCFGs for this task.

(d) Briefly discuss one significant advantage of using SCFGs instead of HMMs for this task.

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) Define collective classification (also known as collective inference).

(b) Describe a concrete task where collective classification would be useful, and discuss why.

(c) How would you implement collective classification for this task? Give details, including your representation, learning algorithm and inference algorithm.

(d) Describe a situation where we would prefer traditional classification learning to the collective version. Justify your answer.

A760-2. For many representations used by learning algorithms, it is possible for some part of a model to be "shared" or "reused." For example, multiple connections in a neural network may share the same weight parameter. For one of the following representations–decision trees, first-order logic programs, Bayesian networks, and stochastic context-free grammars –answer these questions:

(a) Describe one way in which sharing might be entailed in this representation.

(b) Why might the type of sharing used in (a) be beneficial?

(c) How would a standard learning algorithm for this representation be modified to implement sharing?

(d) How would the standard classification procedure be changed, if at all?