SPRING 2003
COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN – MADISON
PH.D. QUALIFYING EXAMINATION

Artificial Intelligence

Friday, January 31, 2003
3:00 - 7:00 PM

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:

- both (2) questions in the section labeled B760, *and*

- any two (2) additional questions in the sections B731, B766, and B776, where these two questions need *not* come from the same section, *and*

- both (2) questions in the section labeled A760.

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 B760. Also answer any two (2) additional questions in any of B731, B766, B776 (these two questions need NOT occur in the same section).

B731 – ADVANCED AI BASIC QUESTIONS

B731-1. Most Bayesian network structure learning algorithms view the task as search.

(a) Define a good search space (states and state-transitions) for this task.

(b) Give two (2) different scoring functions for a state (be precise...you may design your own). List one (1) advantage of each function.

(c) Give two (2) different search strategies that are appropriate for this task. List one (1) advantage and one (1) disadvantage of each.

(d) Suppose the prior probability of the correct Bayesian network having n free parameters is 2^{-n}, while all Bayesian networks with a given number of parameters are equally likely, and all parameter values are equally likely. Modify one of your search algorithms to estimate the posterior probability (given your training data) of the correct Bayesian network having a given arc (say an arc from variable X to variable Y).

B731-2. Consider the relational database below.

(a) Express this relational database in Prolog.

Now suppose each relational table has thousands of tuples (records), including additional customers and products, and you wish to use an ILP system to learn first-order logic rules to predict which customers are likely to buy which products.

(b) Which portions of your Prolog program will serve as positive examples and which will serve as part of your background knowledge?

(c) Do you need negative examples, and if so, how will you get them?

(d) What additional background knowledge would be useful to add?

(e) Give an illustration of the type of clause you hope the ILP system might learn.

Customer

Name	Hobby
Sue	Skiing
George	Ice Fishing

Product

Name	Cost
Ski Poles	$30.00
Ice Saw	$50.00

Purchase

ID	Customer	Product
001	George	Ice Saw
002	Sue	Ski Poles
003	Sue	Ice Saw

B760 – MACHINE LEARNING BASIC QUESTIONS

B760-1. Ensembles are a popular technique for improving supervised learning.

(a) Describe one (1) popular ensemble method and briefly argue why it is an appealing technique.

(b) Some ensemble algorithms use some sort of randomness to encourage diversity among the elements in the ensemble.

(i) Explain what is meant by "diversity" and why it can be important.

(ii) Given a set of training examples and set of hypotheses drawn from some hypothesis space, explicitly define a measure of diversity of the given set of hypotheses.

(iii) Sketch how one might, instead of using randomness to create diversity indirectly, create an algorithm that explicitly tries to maximize diversity during the learning process.

B760-2. If one is not careful designing a machine-learning experiment, the empirical results one obtains can be misleading.

(a) Sketch an improper experimental design for supervised learning, clearly explaining what is wrong with the experiment. (The design error you report needs to be more subtle than simply reporting accuracy on the training data rather than using a separate test set.) Describe how the experiment should instead be done in order to obtain proper results.

(b) Repeat part (a), but this time for reinforcement learning rather than supervised learning.

B766 – COMPUTER VISION BASIC QUESTIONS

B766-1. Consider the use of the *RANSAC robust estimation* method applied to the problem of circle detection (of arbitrary position and size) in a two-dimensional image in which N point features have been detected.

(a) For this application, how many points are used by RANSAC to estimate a candidate circle model?

(b) RANSAC requires that a distance function be defined; define an appropriate distance function for this application, and describe what factors should be used to set the threshold value for use with this function.

(c) Define the "scoring" function that RANSAC uses to determine the goodness of a given candidate circle model.

(d) An alternative scoring function to that used by RANSAC is to compute the median of the distance values from the model to the N data points, and then select the model with the least median value. Compare the use of this scoring function, called LM, with RANSAC's in terms of (i) whether more or fewer parameters need to be set by the user, and (ii) whether one method is more tolerant of a greater percentage of outliers than the other method.

B766-2.

(a) In stereo vision, given a conjugate pair of points, at pixel coordinates $p = (x_l, y_l)$ in the left image, and $q = (x_r, y_r)$ in the right image, give the equation that describes the relationship between these two points when the intrinsic parameters of the two cameras are known but the extrinsic parameters are not known. Include in your answer a description of the properties of each part of your answer.

(b) Define a match function for comparing pixel p and its neighbors in the left image with a pixel q and its neighbors in the right image.

(c) Define another constraint that is useful for determining correct conjugate pairs of pixels. For this constraint, either argue why it is guaranteed to always hold, or else show a scene configuration in which the constraint is violated.

B776 – BIOINFORMATICS BASIC QUESTIONS

B776-1. Consider (soft) EM clustering with Gaussian clusters and (hard) k-means clustering.

(a) Describe the aspects in which they are similar, and the aspects in which they differ.

(b) Discuss one (1) significant advantage of (soft) EM clustering over (hard) k-means.

B776-2. Suppose you are given the task of determining if a newly-discovered document was written by a famous author. For example, the document may be a sonnet potentially authored by Shakespeare. You have at your disposal all of the known writings of the author in question. Suppose you wish to use a Markov model to try to solve this case.

(a) Describe the technical aspects of your model (order, homogenous or inhomogeneous, hidden state or not).

(b) Describe how you will train it.

(c) Describe how you will use it to make the authorship decision.

(d) Briefly discuss one (1) strength and one (1) limitation of your approach.

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. Ima Stu Dent and Soh Am Eye are discussing support vector machines (SVMs). Ima claims SVMs are basically variants of artificial neural networks (ANNs), whereas Soh says SVMs are really much more like *k*-nearest neighbor (*k*-NN) approaches.

(a) What supporting arguments might Ima give for her case? Be sure to discuss neural networks containing hidden units and not just perceptrons.

(b) What supporting arguments might Soh give for his case? Be sure to consider $k > 1$ and not just $k = 1$.

(c) Imagine that you wish to provide more than just labeled input/output pairs to an SVM. Design one (1) possible way of doing so, and mention one (1) strength and one (1) weakness of your approach.

A760-2. In the *active learning* setting, the learning system is given a small set of labeled examples but it can also selectively ask an oracle (e.g. a domain expert) to provide labels for some unlabeled examples. Active learning algorithms typically assume that the learner can ask to have examples labeled one at a time. For some domains, however, this is not a reasonable assumption.

(a) Describe an active learning method in which the learner asks for *sets* of examples to be labeled at one time. Be sure to state the specific learning algorithm (e.g. decision trees) on which your approach is based.

(b) Describe the criteria that the learner should take into account in selecting examples to be labeled by the oracle.

(c) Describe how your approach selects sets of examples for its queries to the oracle.