FALL 2009
COMPUTER SCIENCES DEPARTMENT
UNIVERSITY OF WISCONSIN – MADISON
PH.D. QUALIFYING EXAMINATION

Artificial Intelligence

Monday, September 21, 2009
3:00 – 7:00 p.m.

GENERAL INSTRUCTIONS:

(a) This exam has 19 numbered pages.

(b) Answer each question in a separate book.

(c) 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.*

(d) Return all answer books in the folder provided. Additional answer books are available if needed.

SPECIFIC INSTRUCTIONS:

Answer:

- <u>both</u> questions in the section labeled B760 or B766, corresponding to your chosen focus area, *and*
- any <u>two</u> additional questions in the sections Bxxx, where these two questions need *not* come from the same section, *and*
- <u>both</u> questions in the section labeled A760 or A766, again corresponding to your chosen focus area.

Hence, you are to answer a total of <u>six</u> 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 of the questions in the Section B760 if you are a machine-learning student or both in Section B766 if you are a computer-vision student. In addition, answer two more "B" questions (from any section).

B731 – ADVANCED ARTIFICIAL INTELLIGENCE: BASIC QUESITONS

B731-1. Gibbs Sampling for Bayesian Networks

(a) Describe two different ways Gibbs sampling can be used to address computational issues in Bayesian networks.

(b) Give pseudocode for the Gibbs sampling algorithm for the first of these ways (tasks).

(c) Give pseudocode for the Gibbs sampling algorithm for the second way (task).

(d) What property of a Bayesian network could make Gibbs sampling fail for any one of these tasks, and why? Be specific, using a concrete example if possible.

B731-2. Statistical Relational Learning (SRL)

(a) Define the syntax and semantics of one SRL representation.

(b) Define the syntax and semantics of a second SRL representation.

(c) Describe one advantage of each representation relative to the other.

B760 – MACHINE LEARNING: BASIC QUESTIONS

B760-1: Computational Learning Theory

a) Give an informal definition of Occam's Razor. Absent any other information, would a learning algorithm based on this principle be expected to overfit or underfit data? Briefly justify your answer.

b) Write out the formula for sample complexity in PAC learning. For each variable, write one sentence explaining its meaning. Show where a version of Occam's Razor makes an appearance in this formula and briefly explain its interpretation.

c) What is inductive bias and why is it required in inductive learning?

d) Give an example of a **finite** hypothesis space H on m Boolean valued concepts drawn from noise free training data that is not PAC-learnable; show this by computing the sample complexity using the formula you wrote in (b) above.

e) Consider two hypothesis spaces H_1 and H_2. Let VC be the Vapnik-Chervonenkis dimension. Suppose H_2 is PAC-learnable.

   i. If $\mathrm{VC}(H_1) < \mathrm{VC}(H_2)$, is H_1 always PAC-learnable? Why or why not?

   ii. If $H_1 \subset H_2$, is H_1 always PAC-learnable? Why or why not? (Hint: The answer is no. If you can't think of a counterexample, then simply explain why this is plausible.)

B760-2. Reinforcement Learning.

Consider the task of *reinforcement learning* (RL).

(a) Discuss what *generalizing across state* means in RL. Explain one key strength of doing so and one key weakness.

(b) Another fundamental issue of RL is the exploration/exploitation tradeoff. Draw a three-state RL graph that illustrates the need for *exploration* in *Q*-learning task (name the states *A*, *B*, and *C*). Explain your answer.

(c) Consider some other RL task with states S_1 through S_N and actions A_1 through A_M. Imagine that a human teacher tells your RL algorithm that in State S_2 it is *usually* a good idea to perform this sequence of three actions A_7-A_9-A_{11}. Assume you know that after doing these three actions in State S_2 your RL agent will always end up in S_8. (You should assume the teacher is neither malicious nor mistaken.)

Describe one way to use this teacher-provided hint. Discuss one strength and one weakness (if any) your solution has compared to the case where your RL agent is not given this information.

(d) Repeat Part *c* but this time assume that the teacher says that when in State S_2 the best thing to *always* do is perform actions A_7-A_9-A_{11}. Be sure to discuss how your answer to this part differs from your answer to Part *c*.

B766 – COMPUTER VISION: BASIC QUESTIONS

B766-1. Interest Point Detection

(a) How does the SIFT interest point detector automatically select the best characteristic scale at a point, while also producing a sparse set of points?

(b) Explain why SIFT interest points are often *not* stably detected near contours of regions.

(c) Describe one main reason why matching patches associated with interest points in two different images using normalized cross-correlation may not yield as good correspondence results as comparing SIFT descriptors.

(d) Describe another method for detecting a set of interest/corner points in an image and give one advantage of it over using the SIFT algorithm.

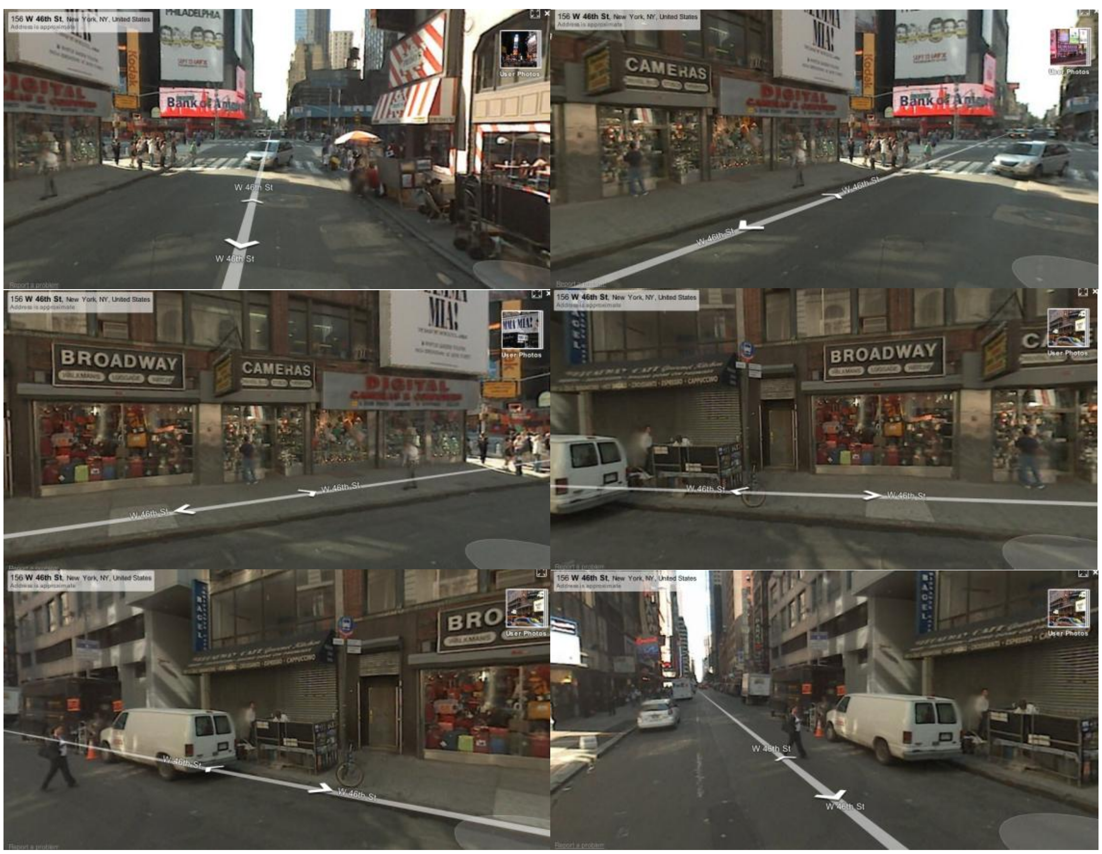
B766-2 Multiple View Stereo

One important topic in computer vision is to extract 3D structure of the world from images.

(a) Consider the following three groups (A,B,C) of pictures downloaded from Google Street View. Among the 3 groups, which group is best suited for reconstructing the 3D structures of the buildings? Which group is impossible? Which group is unclear, depending on some parameters that may be hard to tell by looking at the pictures? For each group, please state the reason (hint: consider motion parallax and whether there is enough information to estimate it).

(b) If we have one camera mounted on a moving car to take this type of picture, we can only reconstruct the street and the buildings but not pedestrians and bicyclists. Why? If we want to reconstruct pedestrians and bicyclists, we need to mount more cameras on the moving car. Suppose you can mount 4 cameras in either of two ways as shown after Group C. Which of these two ways of mounting is better? Explain why.

Group A



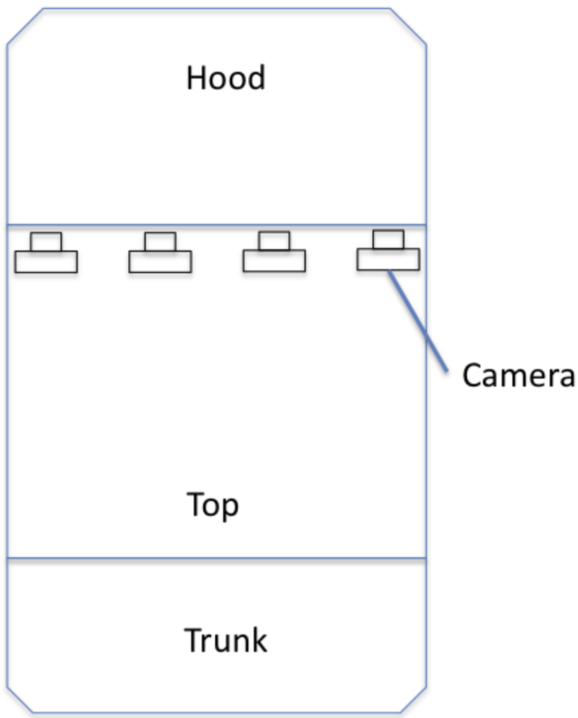
Group B


E Wilson St / W Wilson St / Martin Luther King Jr Blvd
268 Martin Luther King Jr Blvd
258 Martin Luther King Jr Blvd
246 Martin Luther King Jr Blvd
236 Martin Luther King Jr Blvd
226 Martin Luther King Jr Blvd
User Photos

Group C

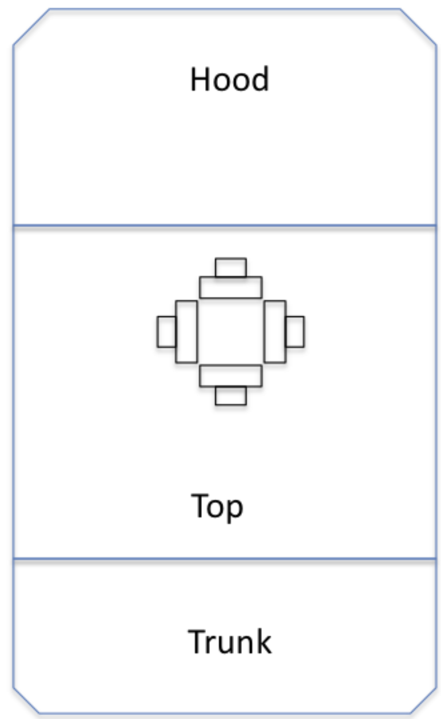

652 W Johnson St, Madison, WI, United States
Address is approximate
User Photos
W Johnson St
W Johnson St
W Johnson St / N Frances St, Madison, WI, United States
Address is approximate
N Frances St
W Johnson St
W Johnson St
N Frances St
578 W Johnson St, Madison, WI, United States
Address is approximate
W Johnson St
W Johnson St
560 W Johnson St, Madison, WI, United States
Address is approximate
W Johnson St
W Johnson St
W Johnson St / Marion St, Madison, WI, United States
Address is approximate
W Johnson St
W Johnson St
Marion St
522 W Johnson St, Madison, WI, United States
Address is approximate
W Johnson St
W Johnson St

Camera installation option 1



A top view of the car

Camera installation option 2



A top view of the car

B769 -- ADVANCED NATURAL LANGUAGE PROCESSING: BASIC QUESTIONS

B769-1. Spectral Clustering

Suppose you have n documents, and you want to group them using spectral clustering.

(a) What is one major advantage of spectral clustering, compared to k-means clustering?

(b) Briefly describe the major steps involved in spectral clustering on your documents.

(c) In general, how many zero eigenvalues does the unnormalized graph Laplacian have? What is the interpretation of their corresponding eigenvectors?

(d) If you know the true cluster labels of a subset of the n documents, how can you incorporate those into spectral clustering? Discuss one approach (hint: think how you may do it with k-means clustering).

B769-2. Hidden Markov Models (HMMs)

La Erton is an extinct language. As the only known written record, linguists have uncovered a La Erton poem with 3,247 word types and 46,392 word tokens. Linguists believe that La Erton has a particularly simple Part-of-Speech (POS) structure, where every word has one of two POS tags: A or B. However, only the first paragraph, or 32 word tokens, of the poem has been POS tagged by linguists to date. The remaining poem is untagged.

(a) Describe how you might build a La Erton POS tagger from the data above, including the untagged part of the poem, using an HMM. Be sure to clearly define the elements of the HMM, and describe their interpretation in this task.

(b) HMM training using the untagged poem has some intrinsic dangers. Describe <u>one</u> hypothetic scenario in which such HMM training will lead to a very poor tagger. Be specific and explain why.

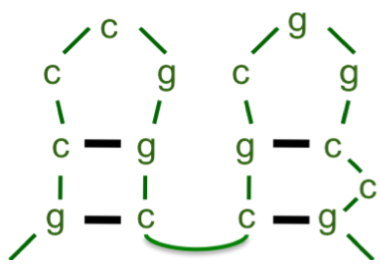
B776 -- ADVANCED BIOINFORMATICS: BASIC QUESTIONS

B776-1. Stochastic Context-Free Grammars

Consider modeling a simple class of RNA secondary structures, as illustrated in the figure below. The thin lines represent the linear sequence of the RNA, and the thick, horizontal lines represent bases that are paired in the secondary structure.

(a) Write down a set of productions and their associated probabilities for this class of RNAs. Assume that the alphabet uses only **c**'s and **g**'s. Your grammar should encode the following requirements

- Both stems always have <u>two</u> sets of paired bases.
- The stems may include a single unpaired base (a "bulge") between the base pairs in the stem (as shown in the right stem).
- The probability of a stem having a bulge is 0.1. The bulge will appear on either side of the stem with equal probability, and it is equally likely to be a **c** or a **g**.
- Each stem position is equally likely to have a **c-g** pairing or a **g-c** pairing.
- Both loops are always three bases long.
- Each loop position is equally likely to have a **c** or a **g**.



(b) Describe how you would modify this grammar to allow loops that have at least <u>three</u> bases but which can be arbitrarily long.

(c) Show an equation that defines $P(l)$, the probability distribution over the lengths (l) of the loops generated by your modified grammar in part (b).

B776-2. Metagenomics Read Clustering

A metagenomics data set consists of DNA-sequence reads from the genomes of many different microbial species that live in a single environment (e.g., water from Lake Monona). A common task with metagenomics data is to cluster reads such that reads in the same cluster are likely to be derived from a single species. This task must be accomplished without knowledge of the species or genomes that may be present in the environment. Computational biologists have found that *sequence composition*, in terms of the set of q-mers (length q substrings) contained within a sequence, can be used for successfully clustering metgenomics reads. You are given a metagenomics data set with N reads, each of length L. Taking advantage of read sequence composition:

(a) Briefly describe how to use a *k-means algorithm* for clustering the metagenomics reads. Assume that k is pre-specified.

(b) Briefly describe how to use an *Expectation-Maximization algorithm* for clustering the metagenomics reads into k clusters (with k pre-specified). Your algorithm must include a *generative* model for the read *sequences*.

Answer both of the questions in the Section A760 if you are a machine-learning student or both in Section A766 if you are a computer-vision student.

A760 – MACHINE LEARNING: ADVANCED QUESTIONS

A760-1. Multiple-Instance Learning and Expectation Maximization (EM)

Consider formulating an *Expectation Maximization*-like approach to learn classification models in a *multiple-instance* setting. You don't need to formally justify that your approach is an actual EM algorithm that will converge, but you should informally explain how it is like an EM algorithm.

(a) Define the multiple-instance learning task.

(b) What do the hidden variables in your approach represent? (Hint: consider what is not known about positive bags).

(c) Now describe your EM-based approach for this task. Be sure to specify:

- the representation used for your models,
- the calculations done in the E-step,
- the calculations done in the M-step,
- any constraints your approach uses to take advantage of what is assumed to hold for positive and negative bags in the multiple-instance setting.

A760-2. LEARNING A GRAPHICAL MODEL

When learning a graphical model, in general one needs to perform *structure learning* and *weight learning*.

(a) Explain what is meant by *weight learning*.

(b) Explain what is meant by *structure learning* and describe how this task can be cast as an AI search problem.

(c) Imagine you choose Markov Logic Networks (MLNs) as your graphical-modeling approach and you choose to use a top-down inductive-logic programming (ILP) to learn the structure of your MLNs.

Assume you wish to use top-down ILP to learn Horn clauses for predicting *p(?x, ?y, John)*, where leading question marks denote universally quantified variables, and you have *q(?x)* and *r(?x, ?y)* as background knowledge. Draw and explain the direct (i.e., immediate) children of the root node in your ILP search space.

(d) Describe <u>one</u> strength and <u>one</u> weakness of using traditional ILP algorithms (such as Progol or Aleph) as the structure learner for MLNs.

(e) Present and justify an approach that overcomes the weakness you identified in Part d.

A766 – COMPUTER VISION: ADVANCED QUESTIONS

A766-1 Image transformation

In computer vision (also in graphics), there are different types of image transformations, e.g., rotation, translation, affine, and homography. One application of these transformations is video stabilization. One simple method is the following: given a shaky video (due to involuntary hand motion), align each frame to its previous frame using these image transformations, so that the resulting video will look much more stable—hence less visually distracting.

(a) Consider the following two situations. For each case, state why or why not the above simple stabilization idea may work effectively to remove the shaky motion in your videos:

i. You stand on a cliff, taking a video with the viewing direction panning across the Grand Canyon.

ii. Mom takes a video of her child while following him walking at home.

(b) To estimate the image transformation between frames, one important technique is to use visual features such as corners. Assume your videos have plenty of corner-like features. Describe one procedure to estimate the inter-frame transformation. Your solution should be robust to foreground target motion.

(c) Unfortunately, not all videos have enough corner-like features, but they may still have sufficient edge features. In this case, discuss how the inter-frame image transformation may be estimated.

A766-2. Graph Partitioning Based Image Segmentation

A popular class of techniques in computer vision views the image segmentation problem as a partitioning problem on an appropriately constructed graph. These methods include both "seeded" algorithms (where some user participation is needed) and techniques that need minimal participation from the user (such as the number of regions desired).

(a) Explain the key steps of graph construction (using a few adjacent pixels) that will permit the use of maximum flow/minimum cut duality to obtain a segmentation of the image into two classes. Include a brief description of the objective function you seek to optimize via this graph.

(b) Assume that the final segmentation determined by your implementation contains a set of small and spurious regions (single pixels assigned to one class surrounded by a large number of pixels from the other class). Describe how you could modify the terms in your objective function to partly mitigate this problem.

(c) Suppose you are also given (by the user) certain must-link constraints such as: pixel *p* must be assigned to class/cluster C1. How can you modify the graph so that your solution satisfies these requirements?

(d) Suppose you are also given information by the user that an adjacent pixel pair (p,q) must be assigned to the same class. How can you modify the graph to enforce this requirement?

(e) Describe one situation where solutions from graph-cut based segmentation may not be well-suited. Explain why.

This page intentionally left blank. You may use it for scratch paper. Please note that this page will NOT be considered during grading.