Artificial Intelligence (4003-455-01)
Introduction to Artificial Intelligence (4005-750-01)

Fall Quarter, 2007 (20071)
Final Examination
Richard Zanibbi

Duration: 120 minutes **Points:** 100 points

Student Name: ____
Section (circle one): Undergraduate Graduate

Instructions:

- This exam is closed book and notes; no electronic devices are not permitted.
- Write your name on the examination booklet(s) that you use during the exam. Hand this sheet and all exam booklets in at the end of the exam.
- **If you do not understand a question or need additional paper, raise your hand.**
- **Make sure to state any assumptions that you make when answering questions.**

Q1 /14	Q2 /14	Q3 /12	Q4 /14	Q5 /16	Q6 /16	Q7 /14			**Total /100**

1. (14 points) **Logical Inference**
   (a) (6) Below is a simple set of horn clauses, given as a PROLOG program. Use backward chaining, matching rules in the order given (as in PROLOG) to prove that Dana is happy. Make sure to indicate the order of steps taken, and unifications made.

$$atrit(dana).$$
$$\neg atrit(rita).$$
$$friend(rita, dana).$$
$$friend(dana, bob).$$
$$friend(X, Y) :- friend(Y, X).$$
$$rich(rita).$$
$$happy(X) :- atrit(X), friend(X, Y), rich(Y).$$

   (b) (2) According the semantics of PROLOG and the program above, is Rita happy? Explain your answer.
   (c) (6) Convert the following statements to conjunctive normal form, and then provide a resolution refutation proof (proof by contradiction using resolution) for the statement that "John is happy" (*happy(John))* (Note: variables are lower case). Use a tree to show the resolution steps, and remember to normalize rules apart and to include all unifications.

$$\forall x \ (pass(x, History) \land win(x, Lottery)) \rightarrow happy(x)$$
$$\forall x \forall y \ (study(x) \land lucky(x)) \rightarrow pass(x, y)$$
$$\neg study(John) \land lucky(John)$$
$$\forall x \ lucky(x) \rightarrow win(x, Lottery)$$

2. (14 points) **Philosophical and Ethical Issues, Future Directions**
   (a) (4) Describe what is meant by *weak* as opposed to *strong* artificial intelligence.
   (b) (6) Identify and briefly discuss two possible negative ramifications if artificial intelligence research is successful in creating a "true" intelligence.
   (c) (4) What is meant by *bounded optimality,* and why is this under active investigation currently, as opposed to *perfect rationality?*

3. (12 points**) Knowledge Representation**
   (a) (4) What is the *qualification problem* for defining actions in rule-based systems? Is there a complete solution to this problem?
   (b) (8) Nada is a student at RIT, and Frieda is a student at the Eastman School of Music. Students normally have one computer, but RIT students normally own two computers *(for the sake of this question).* We also know that Frieda is the proud owner of two bassoons, while students normally own no bassoons.
      i. Draw a semantic network to describe this scenario.
      ii. Using first-order logic, define a predicate that can be used to determine the categories to which Nada and Frieda belong in the semantic network.
      iii. Looking at the network, how does one determine how many bassoons and computers that Frieda and Nada own?

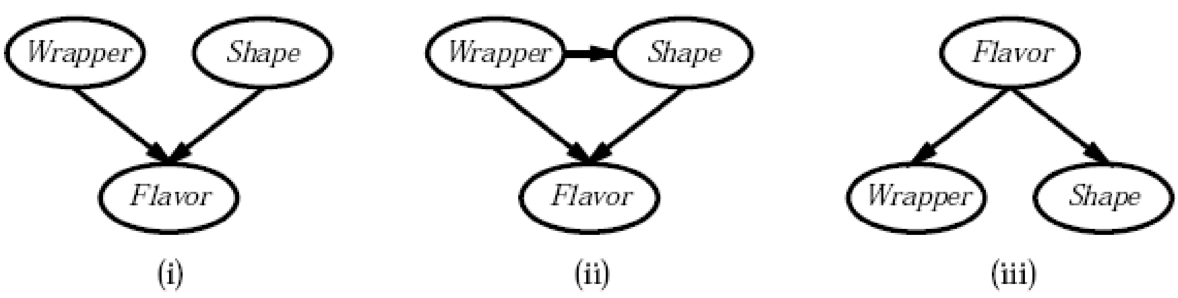
4. (14 points) **Classical Planning**
   (a) (4) How does the representation of states, goals, and actions in planning algorithms differ from that for "standard" search algorithms?
   (b) (5) Provide a solution to the following STRIPS specification of a blocks world problem, in which we want to reverse the blocks in a three block column, with block C on the table, and block A on top. Trace the execution of your solution (including variable bindings): show only the changes after each step. **Note** that *Pickup* removes a block from the table, while *Unstack* removes a block from on top of a block that is *not* on the table, and similar relationships exist for *Putdown* and *Stack. Gripping* indicates which block the robot is holding.

$Init(OnTable(C), OnTop(B, C), OnTop(A, B), Clear(A), Gripping())$
$Goal(OnTable(A), OnTop(B, A), OnTop(C, A))$
Action($Pickup(x)$,
  $PRECOND: Gripping(), Clear(x), OnTable(x)$
  $EFFECT: Gripping(x), \neg OnTable(x), \neg Gripping())$
Action($Putdown(x)$,
  $PRECOND: Gripping(x)$
  $EFFECT: OnTable(x), Gripping(), Clear(x), \neg Gripping(x))$
Action(Stack (x, y)
  $PRECOND: Clear(y), Gripping(x)$
  $EFFECT: OnTop(x, y), Gripping(), Clear(x), \neg Clear(y), \neg Gripping(x))$
Action($Unstack(x, y)$
  $PRECOND: Clear(x), Gripping(), OnTop(x, y)$
  $EFFECT: Gripping(x), Clear(y), \neg Gripping(), \neg OnTop(x, y))$

(c) (4) Describe how a partial-order plan for the planning problem shown in (b) can be constructed. What is the advantage of a partial-order plan over the one that you provided in part (c)?

5. (16 points) **Bayesian Networks**
The Surprise Candy Company makes candy in two flavors: 70% are strawberry and 30% are anchovy flavor. Each new piece of candy starts out with a round shape; as it moves along the production line, a machine randomly selects a certain percentage to be trimmed into a square; then, each piece is wrapped in a wrapper whose color is chosen randomly to be red or brown. 80% of the strawberry candies are round and 80% have a red wrapper, while 90% of the anchovy candies are square and 90% have a brown wrapper. All candies are sold individually in sealed, identical, black boxes.

Now you, the customer, have just bought a Surprise candy at the store but have not yet opened the box. Consider the following three Bayes' nets:



(a) (3) Which network(s) can correctly represent *P(Flavor, Wrapper, Shape)?*
(b) (2) Which network is the best representation for this problem?
(c) (2) True or False: network (i) asserts that *P(Wrapper | Shape) = P(Wrapper).*
(d) (3) What is the probability that your candy has a red wrapper?
i. 0.8 ii. 0.56 iii. 0.59
(e) (3) In the box is a round candy with a red wrapper. The probability that its flavor is strawberry is:

i. <= 0.7 ii. Between 0.7 and 0.99 iii. > 0.99

(f) (3) Explain how conditional probability tables in Bayesian networks may be represented more compactly, similar to how Bayesian networks reduce the representation of joint probability distributions.

6. (16) **Neural Networks and Support Vector Machines**

x1	***x2***	**Expected Output**
2	1	1
0	0	0
-2	3	0

You are training a single perceptron to classify the data above into two classes. The initial weights for the network are [-0.5, 4, 1] for input nodes *x1, x2*, and x3, where *x3* is a bias node, with fixed input value 1. The perceptron learning rule is used for training, with $\alpha = 0.3$ and the step (threshold) function used as the activation function. Note that for the perceptron learning rule, the derivative of the activation function is *not* used to update network weights.

(a) (4) Draw a diagram showing the structure of this neural network, including weights, and define the output of the perceptron in terms of the given network properties.

(b) (6) Show the perceptron output and weight updates after each of the three inputs in the table arrive in order.

(c) **BONUS (+2):** If allowed to iterate over the data, can our perceptron learn to classify the data shown correctly? Explain your answer.

(d) (3) Briefly explain how support vector machines address the problem the problem of classifying data that is not linearly separable. Be sure to explain what the *support* vectors are.

(e) (3) How are inputs classified in a two-class (-1/1 output) support vector machine?

7. (14) **Inductive Learning and Decision Trees**

(a) (2) What does it mean to say that a (computationally) learned hypothesis *generalizes well?*

(b) (4) How is the performance of a learning algorithm normally measured? Make sure to describe how data is used in the process.

(c) (4) Give a decision tree to correctly classify the data shown in question 6.

(d) (4) What is *entropy*, and how is it used in decision tree learning?