CS 188 Fall 1999 — Introduction to AI — Stuart Russell

Midterm Solutions

1. (18 pts.) True/False

(a) (3) False. In inaccessible or stochastic worlds, a rational agent cannot know the outcomes of its actions; moreover, if it's unlucky it will not outperform a nonrational agent.

(b) (3) True. Humans do not always reason soundly, for example.

(c) (3) False. Internal state could be useful for storing the results of computations or storing the results of learning (e.g., a KB).

(d) (3) True. $\{x/F(F(A)), y/F(A), v/F(A)\}$.

(e) (3) False. Can get stuck at local maxima and fail to find a solution.

(f) (3) True. It's also valid.

2. (20 pts.) Search, constraint satisfaction

(a) (5) Initial state: a single square (assuming $n \geq 1$).
Operators: add a single square to an open edge of any existing square.
Goal test: the state contains n squares and passes the beauty test.

(b) (5) No heuristic available, hence uninformed. Goal is at fixed depth, hence depth-limited search (limit = $n-1$) works and uses least space. Might also want repeated-state checking.

(c) (5) The branching factor is the number of open edges on the current state. Adding a new square uses up one open edge and adds three, for a net gain of 2. Hence the maximum branching factor goes like 4, 6, 8, 10 for shapes with 1, 2, 3, 4 squares. The number of shapes of size n is at most the product of the branching factors for shapes up to $n-1$, which is at most $2^{n-1}n!$.

(d) (5) There are at least to ways to do this. (i) Variables are the dominoes; values are *pairs* of adjacent squares on the shape; constraints prevent overlap simply by checking that two pairs of squares are disjoint. (ii) Variables are squares; values are possible adjacent squares that this square will be paired with; constraints say that no two squares have the same value and if A chooses B then B must choose A.

3. (14 pts.) Propositional Logic

(a) (2) 4 variables, hence $2^4 = 16$ models.

(b) (4) An implication is false if the premise is true *and* the conclusion is false. There are 4 models where $R \wedge C$ is true. The negated conclusion is $\neg(\neg O \wedge \neg B)$ which is just $O \vee B$, and this is true in 3 of the 4 models.

(c) (4) Yes. It is equivalent to
$R \wedge C \Rightarrow \neg O$ and $R \wedge C \Rightarrow \neg B$.
In clause form, these become
$\neg R \vee \neg C \vee \neg O$ and $\neg R \vee \neg C \vee \neg B$.
These clause have zero positive literals, and hence are Horn.

(d) (4) To prove that A does not entail B, one simply has to provide a model where A is true and B is false. The model is R, C, $\neg B$, O.

4. (12 pts.) First-order logic
Let $M(x)$ be true if x is a mail carrier; $B(x)$ be true if x lives in Berkeley; and $K(x, y)$ be true if x knows y. Translate the following sentences into first-order logic:

(a) (6) There are at least two mail carriers who live in Berkeley.
$\exists\, x, y \;\; M(x) \wedge M(y) \wedge B(x) \wedge B(y) \wedge x \neq y$

(b) (6) All the mail carriers who live in Berkeley know each other.
$\forall\, x, y \;\; M(x) \wedge M(y) \wedge B(x) \wedge B(y) \;\Rightarrow\; K(x, y)$
(Adding the condition $x \neq y$ is optional but preferred.)

5. (16 pts.) Resolution

(a) (6) Two methods: (i) Show that $A \;\Leftrightarrow\; B$ is valid, by negating it, converting to CNF, proving a contradiction. (ii) Prove that $A \models B$ and $B \models A$, each using the standard procedure. The same work gets done either way.

(b) (10) Converting A to CNF:
$\forall\, x \;\; \neg[\exists\, y \;\; P(x, y)] \vee Q(x)$
$\forall\, x \;\; [\forall\, y \;\; \neg P(x, y)] \vee Q(x)$
A$'$: $\neg P(x, y) \vee Q(x)$
Converting $\neg B$ to CNF:
$\neg[\forall\, x, y \;\; P(x, y) \;\Rightarrow\; Q(x)]$
$\neg[\forall\, x, y \;\; \neg P(x, y) \vee Q(x)]$
$\exists\, x, y \;\; \neg[\neg P(x, y) \vee Q(x)]$
$\exists\, x, y \;\; P(x, y) \wedge \neg Q(x)]$
B1: $P(G, H)$ and B2: $\neg Q(G)$
Resolve A$'$ with B1, $\{x/G, y/H\}$, giving
C: $Q(G)$
Resolving B2 with C gives the empty clause.

6. (20 pts.) Planning

The STRIPS operator $Ride(x, e, f_1, f_2)$ describes the action of a person x riding an elevator e from floor f_1 to floor f_2, and is defined as follows:

(a) (3)

$$Op(\text{ACTION}{:}Call(x, e, f), \text{PRECOND}{:}On(x, f) \wedge On(e, g) \wedge Working(e),$$
$$\text{EFFECT}{:}\neg On(e, g) \wedge On(e, f))$$

(b) (4) One effect axiom states that the person will get to the target floor:
$\forall\, x, e, f_1, f_2, s \;\; On(x, f_1, s) \wedge On(e, f_1, s) \wedge Working(e, s) \;\Rightarrow\; On(x, f_2, Result(Ride(x, e, f_1, f_2), s))$

(c) (4) One possibility is to say that riding the elevator doesn't break it:
$\forall\, x, e, f_1, f_2, s \;\; Working(e, s) \;\Rightarrow\; Working(e, Result(Ride(x, e, f_1, f_2), s))$
Another possibility is to say that people on other floors do not move when someone else rides the elevator.

(d) (3) See figure. Notice that the goal does not include any extraneous conditions.

(e) (3) See figure. The key is that only some of the variables in the step become instantiated.

(f) (3) There are infinitely many ways: Jeb can reach floor 3 via any other sequence of floors, once he has called the elevator. (Note also that the STRIPS formulation we have also allows E to ride Jeb to floor 3, with the same effect.)