CMPSCI 683 Fall 2000

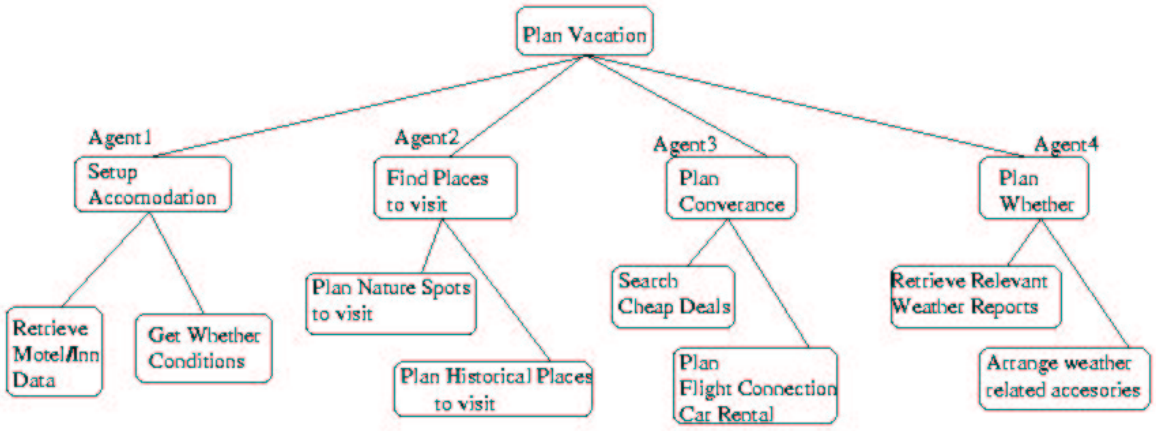
FINAL EXAM Prof. Lesser

1) Agent Architectures/Real-Time (10 points)

(a) Explain how Guardian exploits a separate preprocessor of sensory data to focus computational resources on the key problems that need to be solved, while still permitting important conditions to be responded to quickly.

(b) Do you see any differences in the subsumption architecture of Brooks and the layered hierarchy of modules described in the Question 3 on learning in the robotic soccer application?

2) Multi-Agent Systems (15 points)



Above figure describes a "Plan Vacation" task, there are four agents each works on one subproblem. Consider following issues involved in the agents working together to generate an overall solution:

(a) What type of subproblem interactions will there be between agents? Give examples.

(b) What type of coordination between agents would be worthwhile to solve the interacting subproblems?

(c) What type of changes (if any) would you need to make to your local agent problem-solving strategy in order to coordinate effectively with the other agent?

3) General Learning (15 points)

There has been a Ph.D. thesis on the use of learning in the application of robotic soccer. This is an interesting application because it involves teams of opposing players, the rewards of players? actions are not immediate, the state of all the players is not known to an individual player, the sensing of the environment is noisy and the result of a players action such as kicking are not deterministic.

In order to develop robotic soccer players, the thesis used a learning approach to the development of a player. Instead of trying to perform learning on the entire activities of a player, learned was layered. The learned behavior at one level was used as the basis for learning more complicated behaviors. Learning was used at three levels:

(b) ball-interception (how to capture a moving ball) -- one player

(c) pass evaluation (deciding on the likely success) -- one-one-player

(d) pass selection (deciding who to pass to) -- one-to-many player

1. Explain the potential benefits of this layered learning approach against learning a direct mapping from sensory data to action.

2. Three learning algorithms are used on these three levels respectively: the reinforcement learning algorithm , the neural network learning algorithm and the decision tree algorithm. Which algorithm will you choose for each level ? What reasons would you give for each of these choices about which learning algorithm to use?

3. For learning at level 1 and 2, is instance-based learning appropriate? Explain your answer.

4) Decision Trees/Influence Diagrams/Belief Networks (15)

Recently, researchers have suggested the use of influence diagrams to represent the high-level decision process involved in searching the WWW for information. In order to use influence diagrams in this application they have introduced time and cost of information acquisition into the formula that describes for each node (variable) the value of information gathering with a particular allocation of (time/cost) resources.

K - current information available to the agent

α – *t*he current best decision

$E_{N,j}$ - evidence regarding node (variable) N in the decision model where j=1 ,,, is a possible value

T - given amount of time

C - given cost

O_i - possible outcomes of the user's decision

Q_N (T,C) quality of information obtained

The following formula for the value of information can be formulated

$$V_K(E_N,T,C) = (\sum_j P(E_N = E_{N,j}|K)\, EU(\alpha|K, E_N = E_{N,j}, Q_N(T,C))) - EU(\alpha|K)$$

where $$EU(\alpha|K, E_N, Q_N(T,C)) = \max_A \sum_i P(O_i|K, E_N, Q_N(T, C), Do(A))\, U(O_i, T, C)$$

a) Explain what this formula means (**10**)

b) Given this formula and assuming you can gather one piece of information, what would be the formula that would specify which node to choose next to gather information and how much time and cost you should allocate to this process (**10**).

c) A similar but obviously easier problem occurs in belief networks (i.e., no time or cost considerations) where you need to decide which node to gather information from next. Suppose you had a standard package to evaluate belief networks, how could use the idea of "conditioning," to determine what information to gather next so as to minimize the degree of uncertainty in the network (**7**)?

5) Neural Networks (10 points)

a) How and why are neural networks used as part of a reinforcement learning system?

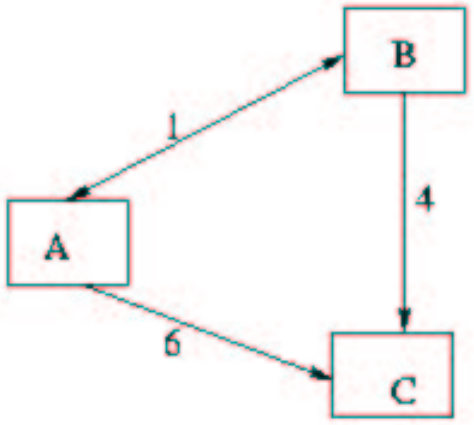
b) How does the number of hidden nodes in a neural network affect its expressability, robustness and learning time?

6) Decision Tree Learning (10 pt)

a) Computer the information gain Gain(A1) and Gain(A2) for the rwo attributes. Show your work. Following information may be of use: I(1/2, 1/2) = 1 bits, I(0,1) = 0 bits, I(1/4, 3/4) = 0.8 bits.

b) Which attribute, A, or A2 would be the better choice? Explain why?

7) Markov Decision Problems (20 pt)



(a) In MDPs, the values of states are related by the following equations, the Bellman equation(P. 503):

$$U(I) = R(I) + \max_a \sum M(a, I, j)\, U(j)$$

where R(I) is the reward associated with being in state I and M(a,I,j) is the probability of reaching state j if action a is executed in state i. Suppose now we wish the reward to depend on actions; i.e. R(a,I) is the reward for doing a in state i. How should the Bellman equation be rewritten to use R(a,I) insteas of R(I)?

(b) Can any finite search problem be translated into a Markov decision problem, such that an optimal solution of the latter is also an optimal solution of the former? If so, explainprecisely how to translate the problem AND how to translate the solution back; if not, explain presicely why not(e.g. give a counterexample).

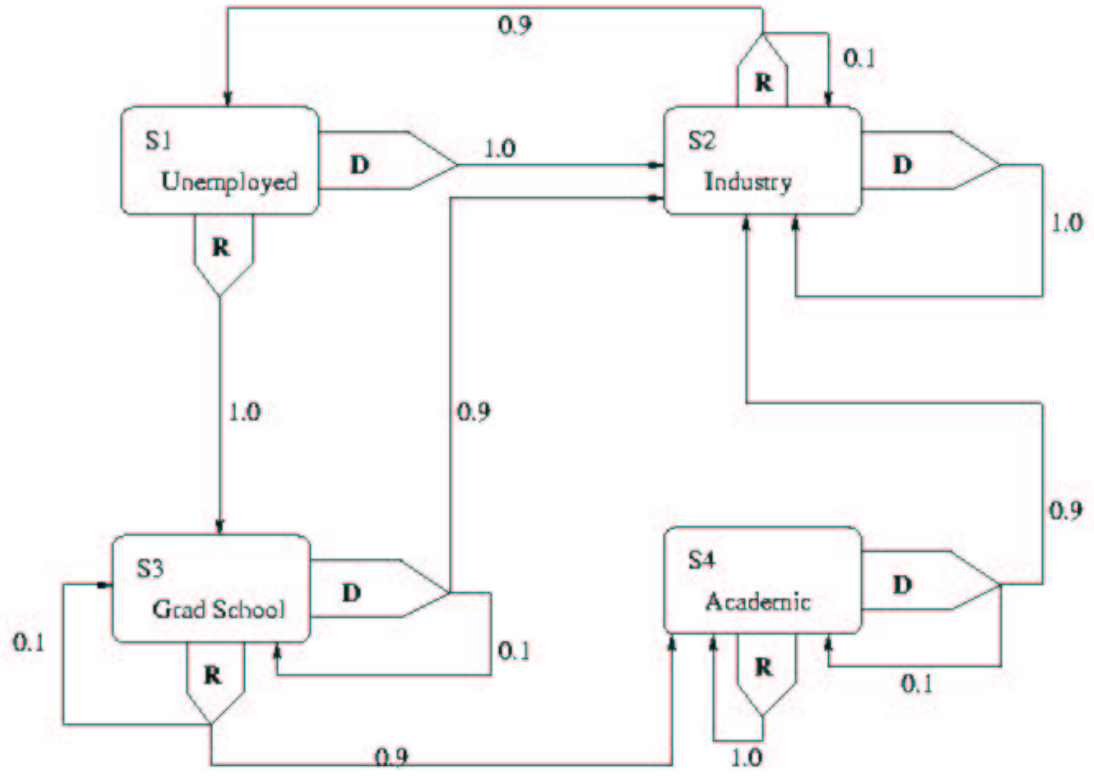
(c) In this part we will apply the value iteration algorith to the MDP that corresponds to the above search problem. Assume that each state has an initial value estimate of 0. Complete the following table, showing the value of each state after each iteration and the optimal action choice given those values. Continue the process until it converges.

State	Iteration							
	0	1	2	3	4	5	6	…
A	0, ->B							
B	0, ->A							
C	0							

(d) Does policy iteration work on this problem? Explain why?

8) Reinforcement Learning (14 points)

Consider the Markov Decision Process below. Actions hava nodeterministic effects, i.e., taking an action in a state returns different states with some probabilities. These transition probabilities are shown in the figure attached to the transition arrows from states and actions to states. There are two actions out of each states: **D** for development and **R** for research.



Consider the following dterministic *ultimately-care -only-about-money* reward for any transition starting at a state*:*

State	S1	S2	S3	S4
Reward	0	100	0	10

1. Let p* be the optimal policy. We allready know, for gamma = 0.9, p*(s)=**D**, for any s=S1, S2, S3, and S4. Compute the optimal values of each state, namely V*(S1), V*(S2), V*(S3) and V*(S4) ACCORDING TO THIS POLICY. *HINT: Start by calcualting V*(S2). Remember the nondeterministic transitions.*
2. According to this same policy, waht is the optimal Q-value, Q(S2,R)?