Exam Markov Decision Theory and Algorithmic Methods (191531920)

January 22, 2016 8:45-11:45h

This exam consists of 4 exercises.
Motivate all your answers.

1. Consider an infinite horizon discounted reward Markov Decision Problem (MDP) with state space $S = \{s_1, s_2\}$, and action sets $A_{s_1} = \{a, b\}$, $A_{s_2} = \{c\}$. The immediate rewards are $r(s_1, a) = 2$, $r(s_1, b) = 6$, $r(s_2, c) = -1$. The transition probabilities are given by $p(s_2|s_1, a) = 1/3$, $p(s_2|s_1, b) = 1$, and $p(s_2|s_2, c) = 1$. The discount factor is $\lambda = 0.90$.

   (a) Use the optimality equations to calculate the discount optimal policy, and the value of this MDP.

   (b) Is the following statement true or false? "The value iteration algorithm may result in an optimal policy."

   (c) Consider the policy iteration algorithm. Let v^n and v^{n+1} be successive values generated by this algorithm. Prove that $v^{n+1} \geq v^n$.

2. The following questions are about small-scale MDPs.

   (a) In an MPD, explain how a policy $\pi \in \Pi^{HR}$ and an initial state s_1 induce a stochastic process of states and actions $(X_1, Y_1, X_2, Y_2, \ldots)$, with X_t the random state and Y_t the random action at time t.

   (b) Consider a finite-horizon MDP. Under which conditions does there exist an optimal deterministic Markovian policy?

   (c) Consider an infinite-horizon average-reward MDP. Using the value iteration algorithm with $\varepsilon > 0$ results in successive values v^n and v^{n+1}. Assume the stop criterion holds. Give two good approximations of the optimal gain g^*.

3. Consider large-scale MDPs with countable state space $S = \{0, 1, \ldots\}$, discount factor λ, and unbounded rewards.

   (a) Let $A_s = \{0, 1, 2, \ldots, M\}$ for all states $s \in S$, $r(s, a) = s$, and $p(j|s, a) = 1$ if $j = s + a$ and $p(j|s, a) = 0$ else, for all $a \in A_s$, $s \in S$. Let $w(s) = \max(s, 1)$, $s \in S$. Show that there exists a constant κ, $0 \leq \kappa < \infty$, such that

$$\sum_{j \in S} p(j|s, a)w(j) \leq \kappa w(s), \qquad \text{for all } a \in A_s, \text{ and all } s \in S.$$

(b) Under suitable conditions (one of them is mentioned in part (a)), the optimality equations have an optimal solution; that is, the MDP has a value. Why does the value iteration algorithm not work in this case?

(c) Explain the method of finite-state approximations. What is the main result of this method?

4. Approximate dynamic programming (ADP) is a recent technique, useful for solving large-scale MDPs.

   (a) Describe and explain the three curses of dimensionality as observed in standard dynamic programming techniques.

   (b) Describe and explain the basic ADP algorithm. How does it address the curses of dimensionality? Mention an advantage and a disadvantage of this algorithm.

Points:

1			2			3			4		Total
a	b	c	a	b	c	a	b	c	a	b	
5	3	4	3	3	3	3	2	3	3	4	+ 4 = 40