






Markov Decision Problems

An Introduction



This Hypercard stack was prepared by:
Dennis L. Bricker,
Dept. of Industrial Engineering,
University of Iowa,
Iowa City, Iowa 52242
e-mail: dbricker@icaen.uiowa.edu

-  Notation & Description
-  "Taxicab" example
-  Linear Programming (LP) Algorithm
-  Policy Improvement (PI) Algorithm
-  More examples

More examples....



Batch processing



Electric generator control



Production planning



Machine replacement



Machine maintenance



Advertising planning



©Dennis Bricker, U. of Iowa, 1998

Given:

- a set S of possible states of the system
- for each state $s_i \in S$, as set K_i of possible actions
- $P_{ij}^k = P\{X_{t+1}=j \mid X_t=i \text{ \& \; action } k \in K_i \text{ is selected}\}$
- $C_{ij}^k = \textit{cost}$ incurred when $X_t=i$, action $k \in K_i$ is selected, and $X_{t+1}=j$



©Dennis Bricker, U. of Iowa, 1998

The Markov Decision Process

At each stage:

- observe the state of the system $i \in S$
- select action $k \in K_i$
- system makes transition to state $j \in S$ with probability P_{ij}^k
- a cost ξ_{ij}^k is incurred or equivalently, an expected cost $C_i^k \equiv \sum_j P_{ij}^k \xi_{ij}^k$ is incurred

©Dennis Bricker, U. of Iowa, 1998

A *policy* R is a vector of actions

$$R = (k_1, k_2, k_3, \dots, k_n)$$

specifying that action $k_i \in K_i$ is to be selected when the system is in state i .

Each policy $R = (k_1, k_2, k_3, \dots, k_n)$ then determines a Markov chain with transition probabilities $P_{ij}^{k_i}$

©Dennis Bricker, U. of Iowa, 1998

Suppose that the Markov chain for every policy has a steady state distribution π^R

What policy R optimizes the expected cost

$$\sum_{i=1}^n \pi_i^R C_i^{k_i}$$

*That is,
which action should be selected in each state in order to optimize the expected cost?*

©Dennis Bricker, U. of Iowa, 1998

Example **Taxi Problem**

A taxi serves 3 adjacent cities: A, B, & C

Each time the taxi discharges a passenger, the

driver must choose from 3 possible actions:

- 1) "cruise" the streets, looking for passenger
- 2) drive to nearest cab stand (hotel, bus station, train station, etc.)
- 3) park and wait for radio call from dispatcher

action 3 is not possible in city B, as distance from transmitter is too great!



©Dennis Bricker, U. of Iowa, 1998

States: $S = \{A, B, C\}$

Actions: $K_A = \{1, 2, 3\}$, $K_B = \{1, 2\}$, $K_C = \{1, 2, 3\}$

state	action	transition probability			reward		
		j=1	2	3	j=1	2	3
A	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	10	4	8
	2	$\frac{1}{16}$	$\frac{3}{4}$	$\frac{3}{16}$	8	2	4
	3	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{5}{8}$	4	6	4

©Dennis Bricker, U. of Iowa, 1998

state	action	transition probability			reward		
		j=1	2	3	j=1	2	3
B	1	$\frac{1}{2}$	0	$\frac{1}{2}$	14	0	18
	2	$\frac{1}{16}$	$\frac{7}{8}$	$\frac{1}{16}$	8	16	8

©Dennis Bricker, U. of Iowa, 1998

state	action	transition probability			reward		
		j=1	2	3	j=1	2	3
C	1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	10	2	8
	2	$\frac{1}{8}$	$\frac{3}{4}$	$\frac{1}{8}$	6	4	2
	3	$\frac{3}{4}$	$\frac{1}{16}$	$\frac{3}{16}$	4	0	8