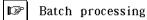


More examples....



Electric generator control

Production planning

Machine replacement

Machine maintenance

Advertising planning



@Dennis Bricker, U. of Iowa, 1998

The Markov Decision Process

At each stage:

- observe the state of the system i ∈ S
- select action k ∈ K_i
- \bullet system makes transition to state j ε S with probability \mathbf{P}_{ij}^k
- \bullet a cost $~\xi_{ij}^k~$ is incurred or equivalently, $~~\text{an expected cost}~~C_i^k \equiv \sum_i p_{ij}^k \, \xi_{ij}^k~\text{is incurred}$

@Dennis Bricker, U. of Iowa, 1998

Suppose that the Markov chain for every policy has a steady state distribution $\pi^{\rm R}$

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?

■ Notation & Description

"Taxicab" example

Linear Programming (LP) Algorithm

Policy Improvement (PI) Algorithm

More examples

@Dennis Bricker, U. of Iowa, 1998

Given:

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



@Dennis Bricker, U. of Iowa, 1998

A policy R is a vector of actions $R = (k_1, k_2, k_3, ... k_n)$ specifying that action $k_i \in K_i$ is to be selected when the system is in state i.

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

@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!

States: $S = \{A, B, C\}$ Actions: $K_A = \{1,2,3\}, K_B = \{1,2\}, K_C = \{1,2,3\}$

te	ion	transition probability			reward		
sta	act	j=1	2	3	j=1	2	3
	1	1/2	1/4	1/4	10	4	8
A	2	1/16	3/4	$3/_{16}$	8	2	4
	3	1/4	1/8	5/8	4	6	4

te	ion	transition probability			reward		
stat	act	j=1	2	3	j= 1	2	3
В	1	1/2	0	1/2	14	0	18
	2	1/16	7/8	1/16	8	16	8

⊚Dennis Bricker, U. of Iowa, 1998

@Dennis Bricker, U. of Iowa, 1998

te	ion	transition probability			reward		
state	act	j=1	2	3	j=1	2	3
	1	1/4	1/4	1/2	10	2	8
С	2	1/8	3/4	1/8	6	4	8 2 8
	3	3/4	1/16	3/16	4	0	8

⊚Dennis Bricker, U. of Iowa, 1998