Exercitii set 1

Luciana Morogan

Academia Tehnica Militara

May 25, 2025

Cuprins

Intr-un enviroment oarecare se considera cunoscute recompensele pentru fiecare stare presupusa a MRP-ului. La time-step-ul t se cunoaste

$$G_t = R_{t+1} + \gamma R_{t+2} + ... = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$$

Se cere mentionarea valorii corespunzatoare obtinerii lui R din secventa G_t la time-step-ul $k+1 = 3$, considerand discountul $\gamma = 0.5$ si $R = +2$.

Intr-un enviroment oarecare se considera cunoscute recompensele pentru fiecare stare presupusa a MRP-ului. La time-step-ul t se cunoaste

$$G_t = R_{t+1} + \gamma R_{t+2} + ... = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}$$

Se cere mentionarea valorii corespunzatoare obtinerii lui R din secventa G_t la time-step $k + 1 = 3$ considerand discountul $\gamma = 0.5$ si $R = +2$

Raspuns. $\gamma^k R = 0.5^2 * 2 = 0.5$

Adevarat sau fals? Argumentati

Intr-un proces de decizie Markov, un factor de reducere γ mare reprezinta faptul ca recompensele pe termen scurt sunt mult mai influente decat recompensele pe termen lung.

Adevarat sau fals? Argumentati

Intr-un proces de decizie Markov, un factor de reducere γ mare reprezinta faptul ca recompensele pe termen scurt sunt mult mai influente decat recompensele pe termen lung.

Raspuns. Fals

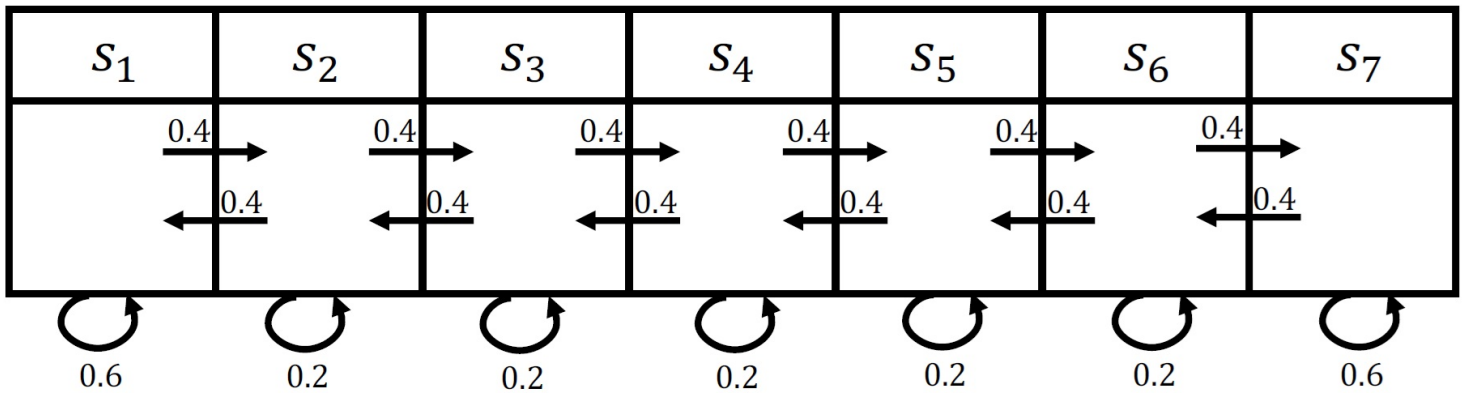
Fie modelul MDP-ul din figura

- Multimea starilor: Locația roverului $(s_1, ...s_7)$
- Acțiunile (tranzitiile dintre stari): *Stanga* sau *Dreapta*
- Recompensele:
  - $+1$ în starea s_1
  - $+10$ în starea s_7
  - 0 în toate celelalte stări

Tranzitiile / dinamica modelului prezice urmatoarea stare a agentului: $p(s_{t+1} = s' | s_t = s; a_t = a)$

Modelul de recompensa prezice recompensa imediata:

$$r(s_t = s; a_t = a) = E[r_t | s_t = s; a_t = a]$$



3. 1.

Daca presupunem $\pi(s_1) = \pi(s_2) = ... = \pi(s_7) = \textit{Dreapta}$, atunci specificati daca aceasta politica este stocastica sau determinista. Justificati.

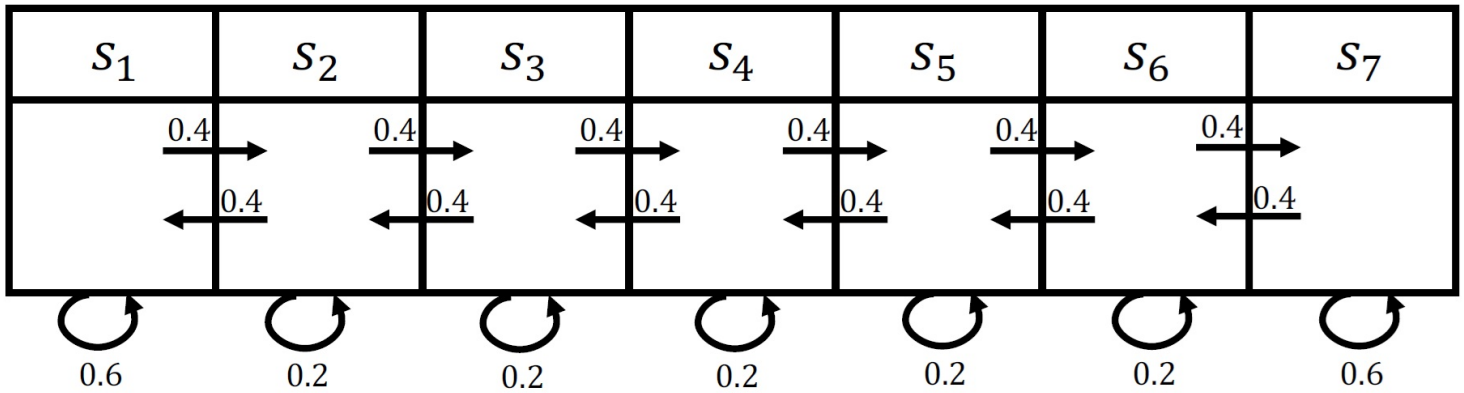
Daca presupunem $\pi(s_1) = \pi(s_2) = ... = \pi(s_7) = \textit{Dreapta}$, atunci specificati daca aceasta politica este stocastica sau determinista. Justificati.

Raspuns. Determinista

3. 2.

- Pentru procesul Markov considerat, Scrieti matricea probabilitatilor de tranzitie a starilor.

- Pentru procesul Markov considerat in figura de mai jos, Scrieti matricea probabilitatilor de tranzitie a starilor.



$$P = \begin{pmatrix} 0.6 & 0.4 & 0 & 0 & 0 & 0 & 0 \\ 0.4 & 0.2 & 0.4 & 0 & 0 & 0 & 0 \\ 0 & 0.4 & 0.2 & 0.4 & 0 & 0 & 0 \\ 0 & 0 & 0.4 & 0.2 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0.4 & 0.2 & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 0.4 & 0.2 & 0.4 \\ 0 & 0 & 0 & 0 & 0 & 0.4 & 0.6 \end{pmatrix}$$

3. 4.

- Pentru procesul Markov considerat anterior, descrieti trei episoade ale lantului Markov pornind din S_4

Pentru:

- dinamica $p(s_6|s_6, a_1) = 0.5, p(s_7|s_6, a_1) = 0.5...$
- recompensele: pentru toate actiunile, $+1$ in starea s_1, $+10$ in starea s_7, 0 altfel

Fie $\pi(s) = a_1, \forall s$
Consideram $V_k = [1\ 0\ 0\ 0\ 0\ 0\ 10]$ si $k = 1, \gamma = 0.5$
Calculati $V_{k+1}(s_6)$

Pentru:

- dinamica $p(s_6|s_6, a_1) = 0.5, p(s_7|s_6, a_1) = 0.5...$
- recompensele: pentru toate actiunile, $+1$ in starea s_1, $+10$ in starea s_7, 0 altfel

Fie $\pi(s) = a_1, \forall s$
Consideram $V_k = [1\ 0\ 0\ 0\ 0\ 0\ 10]$ si $k = 1, \gamma = 0.5$
Calculati $V_{k+1}(s_6)$

Raspuns.

$$
\begin{aligned}
V_{k+1}(s_6) &= r(s_6) + \gamma \sum_{s'} p(s'|s_6, a_1) V_k(s') && (1)\\
&= 0 + 0.5 * (0.5 * 10 + 0.5 * 0) && (2)\\
&= 2.5 && (3)
\end{aligned}
$$