An Introduction to Markov Decision Process

Part 2

Alireza Kavoosi
School of Industrial Engineering, University of Tehran

March 8, 2025

Overview

1. Value Iteration Algorithm

2. Policy Iteration Algorithm

3. Hybrid Value/Policy Iteration Algorithm

Value Iteration Algorithm

[H] **Input:** Tolerance parameter $\epsilon > 0$

Output: Value function v^ϵ and policy π^ϵ

Initialize $v^0(s) = 0$ for all $s \in \mathcal{S}$ Set $n = 1$

repeat

  foreach $s \in \mathcal{S}$ **do**

    $$v^n(s) = \max_{a \in \mathcal{A}} \left(C(s,a) + \gamma \sum_{s' \in \mathcal{S}} \mathbb{P}(s'|s,a) v^{n-1}(s') \right)$$

until $\|v^n - v^{n-1}\| < \frac{\epsilon(1-\gamma)}{2\gamma}$;

return $v^\epsilon = v^n$ and the policy π^ϵ

Policy Iteration Algorithm

[H] **Input:** Initial policy π^0

Output: Optimal policy a^*

Set $n = 1$

repeat

  Compute transition matrix $P^{\pi^{n-1}}$ Compute contribution vector $c^{\pi^{n-1}}(s) = C(s, A^{\pi^{n-1}})$ Solve $(I - \gamma P^{\pi^{n-1}})v = c^{\pi^{n-1}}$ for v^{π^n}

  foreach $s \in \mathcal{S}$ **do**

    $a^n(s) = \arg\max_{a \in \mathcal{A}} \left(C(a) + \gamma P^{\pi^n} v^n \right)$

until $a^n(s) = a^{n-1}(s)$ *for all* s;

return $a^* = a^n$

Hybrid Value/Policy Iteration Algorithm

[H] **Input:** Tolerance parameter ϵ, inner iteration limit M
Output: Policy a^ϵ

Set $n = 1$, choose initial value function $v^0 \in \mathcal{V}$
repeat
  foreach $s \in \mathcal{S}$ **do**
    $$a^n(s) = \arg\max_{a \in \mathcal{A}} \left(C(s,a) + \gamma \sum_{s' \in \mathcal{S}} \mathbb{P}(s'|s,a) v^{n-1}(s') \right)$$
  Set π^n accordingly
  Partial Policy Evaluation:
  Set $m = 0$, $u^n(0) = c^\pi + \gamma P^{\pi^n} v^{n-1}$
  if $\|u^n(0) - v^{n-1}\| < \frac{\epsilon(1-\gamma)}{2\gamma}$ **then**
    Go to Step 3
  else
    for $m < M$ **do**
      $u^n(m+1) = c^\pi + \gamma P^{\pi^n} u^n(m)$ Set $m = m + 1$
  Set $v^n = u^n(M)$, increment n, and return to Step 1
until *convergence*;
return $a^\epsilon = a^{n+1}$

Thank you for your attention!