Ma 661 Dynamic Programming and Reinforcement Learning

Darinka Dentcheva darinka.dentcheva@stevens.edu

Homework 3

Solutions

Consider the equipment location example; its text is repeated for convenience. A service man moves between 4 sites, with site 1 denoting home office, and 2, 3, and 4 denoting remote sites. Work at sites 2,3, and 4 requires the use of an equipment trailer. The cost of relocating the equipment trailer between the sites is $d(j,k) = 300$ for $k \neq j$. The cost $c(k,j)$ of using the trailer is 100 if the work is at site $k > 1$ and trailer is at site $j \neq k$ with $j > 1$; 50 if $j = k$ and $j > 1$, and 200 if the work is at $k > 1$ and the trailer is at $j = 1$ (home office). If the service man is at the home office, no work is done and cost is zero. At any time, the service man knows their location, observes the location of the next job (or 1 if no job is to be done) and decides whether and where to move the trailer. The transition matrix between job locations is

$$P = \begin{bmatrix} 0.1 & 0.3 & 0.3 & 0.3 \\ 0.0 & 0.5 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.8 & 0.2 \\ 0.4 & 0.0 & 0.0 & 0.6 \end{bmatrix}$$

For example, if the current location (job) is 2, the probability that the next job is at 3 is 0.5. Assume the discount factor of 0.95.

(a) Formulate a Markov decision problem to help the repairment decide on the movement of the trailer: define the state and control space and write the dynamic programming equations.

The solution is provided in a separate file.

(b) Solve these equations by the value iteration method (starting from 0) and calculate lower and upper bound at each step.

Solution: Value iteration:

$$v^{i+1}(j,k) = \min_{u \in \{1,2,3,4\}} \left\{ C(j,k,u) + \alpha \mathbb{E}_{P[k,\cdot]} v^i(u,\cdot) \right\}, \forall j,k.$$

with $v^0(j,k) = 0, \forall j,k.$
We get

$$v^* = \begin{bmatrix} 1497.76402403601 & 1618.05406445601 & 1546.27028176717 & 1591.73334413677 \\ 1428.21823566599 & 1494.24454064649 & 1546.27028176717 & 1494.70448733274 \\ 1220.43065205501 & 1318.05406445601 & 1246.27028176717 & 1311.07825065327 \\ 1330.11931047055 & 1492.68995482812 & 1439.28889744161 & 1291.73334413677 \end{bmatrix},$$

in 669 iterations.

As we learn in class, with $\alpha = 0.95$, the value iteration method is slow. So we construct respectively

$$\begin{cases} v^i_{\text{lo}}(j,k) = v^i_{\text{lo}}(j,k) + \dfrac{\alpha}{1-\alpha}\inf\{v^i - v^{i-1}\}, \\ v^i_{\text{up}}(j,k) = v^i_{\text{up}}(j,k) + \dfrac{\alpha}{1-\alpha}\sup\{v^i - v^{i-1}\}, \end{cases}$$

and as

$$v^i_{\text{lo}} \leq v^* \leq v^i_{\text{up}} \leq v^i, \forall i,$$

v^i_{up} converges to v^* faster than v^i, with the stopping criterion $v^i_{\text{lo}} \approx v^i_{\text{up}}$.

Numerically we get the same result with 639 iterations.

(c) Solve the problem by the policy iteration method starting from "do not move trailer from base" at each state.

Solution: Policy iteration:

- v^i is the solution of

$$v^i(j,k) = C(j,k,\pi^i(j,k)) + \alpha \mathbb{E}_{P[k,\cdot]} v^i(\pi^i(j,k),\cdot), \forall j,k.$$

- For all j, k,

$$\pi^{i+1}(j,k) = \arg\min_{u\in\{1,2,3,4\}} \left\{C(j,k,u) + \alpha \mathbb{E}_{P[k,\cdot]} v^i(u,\cdot)\right\}$$

- Initially

$$\pi^0(j,k) = j, \forall j,k.$$

We get

$$v^* = \begin{bmatrix} 1497.76402403601 & 1618.05406445601 & 1546.27028176717 & 1591.73334413677 \\ 1428.21823566599 & 1494.24454064649 & 1546.27028176717 & 1494.70448733274 \\ 1220.43065205501 & 1318.05406445601 & 1246.27028176717 & 1311.07825065327 \\ 1330.11931047055 & 1492.68995482812 & 1439.28889744161 & 1291.73334413677 \end{bmatrix},$$

and

$$\pi^* = \begin{bmatrix} 1 & 3 & 3 & 4 \\ 2 & 2 & 3 & 2 \\ 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 \end{bmatrix},$$

in only 2 iterations.

(d) Solve the problem by linear programming.

Solution: According to the class, v^* is also the optimal solution of the following LP problem:

$$\begin{aligned} \max \quad & \sum_{j,k} v(j,k), \\ \text{s.t} \quad & v(j,k) \leq C(j,k,u) + \alpha \mathbb{E}_{P[k,\cdot]} v(u,\cdot), \forall j,k,u = 1 \cdots 4. \end{aligned}$$

And we get the same numerical result for v^*.