Assignment on Large-Scale Markov Decision Problems, December 2018

Part of the course Markov Decision Theory and Algorithmic Methods (191531920).

Goal

To learn to use algorithms and methods from approximate dynamic programming to solve a large-scale MDP.

Problem: Nomadic trucker project

Assume a trucker may visit any of 20 cities in a set I. The trucker may move loaded from i to j earning a positive reward r_{ij} as long as there is a demand D_{tij} at time t. Alternatively, the trucker may move empty from i to j at a cost c_{ij}.

Assume that the random demands $\hat{D}_{tij}$ (realisations d_{tij}) follow a Poisson distribution with mean λ_{ij}; they are independent over time. Randomly generate these means by first generating a parameter ρ_i for each i, where ρ_i is uniformly distributed between 0 and 1. Then set $\lambda_{ij} = 2\rho_i(1-\rho_j)$. Let m_{ij} be the distance between i and j, with $m_{ii} = 0$. Randomly generate distances from a uniform distribution between 100 and 1500 miles. Now let $r_{ij} = \rho_i m_{ij}$ and $c_{ij} = 1.2 m_{ij}$. Assume that if our trucker is in location i, he can only serve demands out of location i, and that any demands not served at one point in time are lost. Further assume that it takes one day (with one time period per day) to get from location i to location j (regardless of the distance). We wish to solve our problem over a horizon $T = 21$ days.

At location i, the trucker may choose to move loaded to j if $d_{tij} > 0$, or move empty to j. Let $x^L_{tij} = 1$ if he moves loaded from i to j on day t, and 0 otherwise. Similar, let $x^E_{tij} = 1$ if he moves empty from i to j on day t, and 0 otherwise. Of course, $\sum_j (x^L_{tij} + x^E_{tij}) = 1$. We make the decision by solving

$$\max_x \sum_j \left((r_{ij} + \bar{V}^{n-1}(j)) x^L_{tij} + (-c_{ij} + \bar{V}^{n-1}(j)) x^E_{tij} \right)$$

subject to the constraint that $x^L_{tij} = 0$ if $d_{tij} = 0$. (This is inspired by problem 4.18 of Powell.)

(a) Assume our trucker starts in city 1. Use approximate value iteration with a stepsize of $\alpha_{n-1} = 10/(9+n)$. (You do not have to use this algorithm but you may choose another ADP algorithm. Motivate your choice.) Train the value functions for 1000 iterations. Then, holding the value functions constant, perform an additional 1000 simulations, and report the mean and standard deviation of the profits, as well as the number of times the trucker visits each city. What is a good choice for the initial value function?

Extra questions:

(b) Repeat part (a), but this time insert a loop over all cities, so that for each iteration n and time t, we pretend that we are visiting every city to update the value of being in the city. Again perform 1000 iterations to estimate the value function, and then perform 1000 testing iterations.

(c) Repeat part (a), but this time, after solving the decision problem for location i and updating the value function, randomly choose the next city to visit.

(d) Compare your results in terms of solution quality and computational requirements (measured by how many times you solve a decision problem).

Instructions

1. You should solve part (a), and you may solve parts (b)-(d) for a bonus.
2. You may work on this assignment in pairs; for example, you may work with the same partner as for the small-scale MDP assignment. You may of course discuss the assignment with others, but each group should hand in its own work. So, do not copy the solutions of your fellow students!
3. Prepare a report, containing at least the generated data, the used algorithm, the results, and a discussion of the results. Put your code in the appendix. This report may be informal.
4. Share the work equally among the group members. Plan your work because of time constraints.

Result

The report should be handed in via Canvas not later than Friday February 1, 2019.

Grading

The grade for this work is 10% of your final grade for this course.