Aalto University
School of Science

Stochastic shortest path problem

Theory and value iteration

Alvar Kallio
Presentation *#19*
13.11.2020

MS-E2191 Graduate Seminar on Operations Research
Fall 2020

Outline

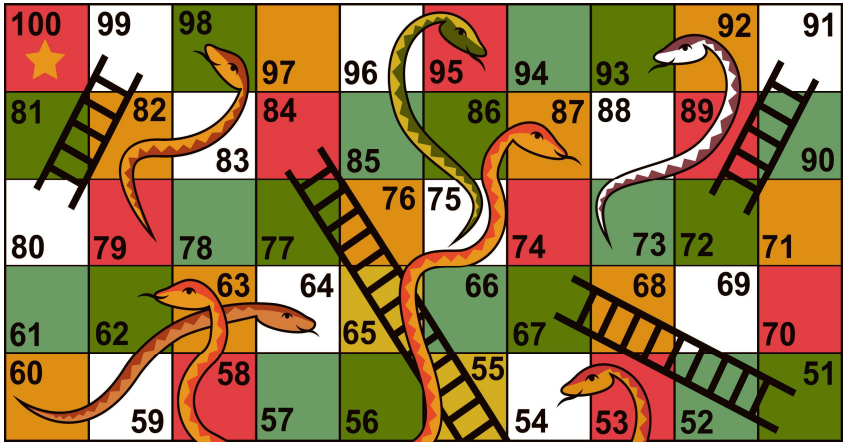
- Example problems and formulation
- Similarities to earlier presentations
- Value Iteration
- Homework and questions

Example 1: Pizza delivery

- Recall the simple example of food delivery
- The courier must find the shortest path
- Transitions are uncertain, the courier chooses the direction
- In this example, the costs do not depend on the chosen direction or the current state

Example 2: Snakes and Ladders

- Recall the Snakes and Ladders game
- Assume that you can either throw a die with 4 or 6 sides
- Which die should you choose in each square?



https://medium.com/re-form/the-timelessness-of-snakes-and-ladders-4ae7d205a4e7

Problem formulation

- States $\{1, \dots, n, t\}$. State t is the destination state.
- At state i a control $u \in U(i)$ is chosen
- The expected cost of choosing control u at state i is $g(u, i)$
- When control u is chosen at state i, the probability of moving to state j is $p_{ij}(u)$
- For convenience, $p_{tt}(u) = 1$ and $g(u, t) = 0$ for any $u \in U(t)$

Problem formulation

- The goal is to find a policy μ describing the optimal actions at each state
- If the transition matrix related to the policy is P_μ and the costs per stage are g_μ the decision maker minimizes the expected cost

$$J_\mu = \sum_{k=0}^{\infty} P_\mu^k g_\mu$$

Connections to earlier presentations

Connection to the deterministic problem

- If the transition probabilities are all either ones or zeros, we get the deterministic shortest path problem
- The costs $g(s,u)$ correspond to the arc weights and there is an arc between nodes i and j if there is an action such that $p_{ij}(u) = 1$
- Solving the problem with value iteration is equivalent to updating labels in presentation 10 repeatedly

Proper and improper policies

- **A policy is called proper if**
  - The target state is reached eventually with probability 1 $\Leftrightarrow$
  - Regardless of the current state, there is a positive probability to hit the destination state at most n steps

- **To get a well formulated problem, we assume that**
  - There exists at least one proper policy
  - For every improper policy, the cost diverges to ∞ for some initial state

Stochastic shortest path problem as a Markov Decision Process

- The problem can be viewed as a MDP as we have a state space, action space, transition probabilities $P(s, u, s')$ and costs $g(s, u)$ (Presentation 5)

- The problem can thus be solved using value iteration, policy iteration or linear optimization as we have seen before (Presentations 12 and 13)

Connection to infinite horizon problems

- This problem is a special case of infinite horizon problems (Presentation 11) with a discount factor $\alpha = 1$
- Similar notation is used, and similar results hold (such as the Bellman Equation)
- The Value iteration method is a variant of the DP algorithm

Value iteration

The cost vector is updated until the results are satisfying

Algorithm 1: Value iteration

Initialize J_0 as you wish;
$n \leftarrow 0$;
while *Termination condition not satisfied* **do**
  for $i \in S$ **do**
    $\mu^*(i) = \text{argmin}_u \, g(i,u) + \sum_j p_{ij}(u) J_n(j)$;
    $J_{n+1}(i) = \min_u \, g(i,u) + \sum_j p_{ij}(u) J_n(j)$;
  end
  $n \leftarrow n + 1$;
end
Return μ^*, J_{n+1};

References

Bertsekas, D. P. (2012). Dynamic programming and optimal control (Vol. 2, 4th ed.) Approximate Dynamic Programming. Belmont, MA: Athena scientific. (pp. 172-189)

Earlier presentations in this course

Homework

The given template solves the example problem *Snakes and Ladders.* Explore how the solution changes when the game board or the used dice are changed or do any other experiment you want. Write a **short** description of your experiment and results.

mail: alvar.kallio@aalto.fi