AI Planning
Lecture 9

Rostislav Horčík

Czech Technical University in Prague
Faculty of Electrical Engineering
xhorcik@fel.cvut.cz

Stochastic Shortest Path

Stochastic Shortest Path

Definition

A stochastic shortest path problem (SSP) is a tuple $\Sigma = \langle S, A, T, s_0, G \rangle$ where

- S is a finite set of states,
- A is a finite set of actions associated with a strictly positive cost function $\text{cost} : A \to \mathbb{R}^+$,
- $s_0 \in S$ is an initial state,
- $G \subseteq S$ is a set of goal states,
- $T : S \times A \to 2^S$ is a nondeterministic transition function assigning to a state-action pair $\langle s, a \rangle$ a set of possible successor $T(s, a)$. If $T(s, a) \neq \emptyset$, there is a probability $P(s' \mid s, a)$ for $s' \in T(s, a)$.

Terminology and assumptions

- We assume that $T(s, a) = \emptyset$ for each goal state $s \in G$ and $a \in A$.
- A state $s \in S \setminus G$ such that $T(s, a) = \emptyset$ for all $a \in A$ is called an deadend.
- We assume that our SSPs have no deadends.
- If $T(s, a) \neq \emptyset$, we say that the action a is applicable in s.
- The set of applicable actions in s is denoted $A(s)$.

Policy

Definition

A map $\pi : S \setminus G \to A$ is called a policy if $\pi(s)$ is applicable in s for each state $s \in S \setminus G$.

Further, we say that $S' \subseteq S$ is closed with respect to π if

- $s_0 \in S'$,
- $T(s, \pi(s)) \subseteq S'$ for each $s \in S'$.

Least closed subset

Given a policy π and a state s, $\widehat{T}(s, \pi)$ is the set of reachable states from s following π, i.e.,

1. $s \in \widehat{T}(s, \pi)$,
2. if $t \in \widehat{T}(s, \pi)$ and $t' \in T(t, \pi(t))$, then $t' \in \widehat{T}(s, \pi)$.

Lemma

$\widehat{T}(s_0, \pi)$ is the least closed subset of S containing s_0.

Proper policy

Definition

A policy π is called proper if following it reaches a goal state with probability 1. We will consider only SSPs having at least one proper policy.

Value function

Definition

Let π be a proper policy. Its value function $V^\pi : S \to \mathbb{R}_0^+$ assigns to a state s the expected cost of reaching a goal state starting in s and following π. It satisfies the following recursive equation:

$$V^\pi(s) = \begin{cases} 0 & \text{if } s \in G, \\ \text{cost}(\pi(s)) + \sum_{t \in T(s,\pi(s))} P(t \mid s, \pi(s)) \cdot V^\pi(t) & \text{otherwise.} \end{cases}$$

Optimal policy

Definition

The optimal policy π^* has the minimum expected cost $V^*(s)$ of reaching a goal from any state s. Its value function $V^*(s)$ satisfies Bellman equation:

$$V^*(s) = \begin{cases} 0 & \text{if } s \in G, \\ \min_{a \in A(s)} \text{cost}(a) + \sum_{t \in T(s,a)} P(t \mid s, a) \cdot V^*(t) & \text{otherwise.} \end{cases}$$

Greedy policy

Definition

Given a function $V : S \to \mathbb{R}_0^+$, we can define a greedy policy π^V with respect to V by

$$Q(s,a) = \text{cost}(a) + \sum_{t \in T(s,a)} P(t \mid s,a) \cdot V(t).$$

$$\pi^V(s) = \operatorname*{argmin}_{a \in A(s)} Q(s,a)$$

Bellman operator

Definition

Let $\mathcal{V}$ be the set of functions $V : S \to \mathbb{R}_0^+$ such that $V(s) = 0$ for each $s \in G$.

An operator $\mathcal{B} : \mathcal{V} \to \mathcal{V}$ defined by $\mathcal{B}(V) = V'$ where

$$V'(s) = \min_{a \in A(s)} \text{cost}(a) + \sum_{t \in T(s,a)} P(t \mid s, a) \cdot V(t)$$

for $s \in S \setminus G$ is called Bellman update (backup) operator.

Value iteration

The optimal value function V^* is the fixed point of $\mathcal{B}$, i.e., $\mathcal{B}(V^*) = V^*$.

Its computation is the algorithm known as Value Iteration (VI). It starts with a function $V_0 \in \mathcal{V}$ and computes successively

$$V_0, V_1 = \mathcal{B}(V_0), V_2 = \mathcal{B}(V_1), \dots$$

Theorem

We have $\lim_{n\to\infty} V_n = V^*$.

In practice, the termination criterion is given by checking whether the residual $\mathrm{Res}_i = \max_{s\in S} |V_i(s) - V_{i+1}(s)| \leq \varepsilon$ for a chosen $\varepsilon > 0$.

Heuristic search

Heuristic

Definition

A heuristic $h\colon S \to \mathbb{R}_0^+$ is called

- admissible if $h(s) \leq V^*(s)$ for all $s \in S$,
- goal-aware if $h(s) = 0$ for all $s \in G$ (i.e., $h \in \mathcal{V}$),
- consistent if $h \leq \mathcal{B}(h)$.

Lemma

Goal-awareness and consistency implies admissibility.

$$h \leq \mathcal{B}(h) \leq \mathcal{B}^2(h) \leq \lim_{n\to\infty} \mathcal{B}^n(h) = V^*.$$

Heuristic search

Running VI on the entire state space is computationally demanding.

Heuristic search works with smaller parts of the state space. It is based on the following steps:

1. an admissible heuristic provides an initialization for the value function,
2. expanding a part of the state space,
3. running VI on a selected subset of states.

ILAO*

Require: SSP $\Sigma = \langle S, A, T, s_0, G \rangle$, an admissible heuristic h, $\varepsilon > 0$

$V \leftarrow h$

$\mathrm{Res} \leftarrow \infty$

while $\mathrm{Res} > \varepsilon$ **do**

Build the closed subset $S' \subset S$ following π^V

while $\mathrm{Res} > \varepsilon$ and $\pi^V = \pi^{V'}$ **do**

$V' \leftarrow V$

$V \leftarrow \mathcal{B}_{S'}(V)$

$\mathrm{Res} \leftarrow \max_{s \in S'} |V(s) - V'(s)|$

return π^V