Analyse eines Forschungsthemas
Stochastic Shortest Paths

Maximilian Starke

Fakultät für Informatik
Technische Universität Dresden

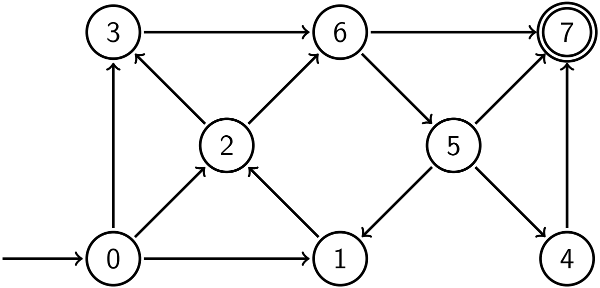
5. September 2022

Inhalt

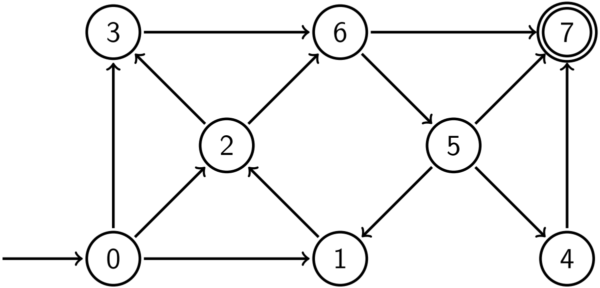
Section 1

Introduction

- The *simplest* shortest path problem

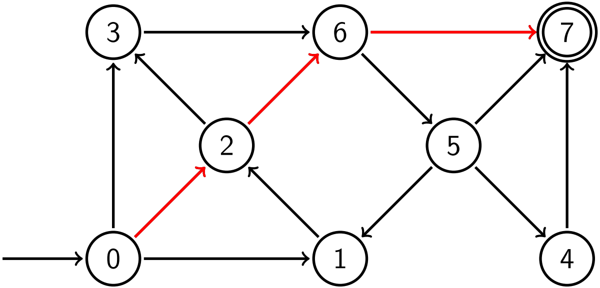


- The *simplest* shortest path problem



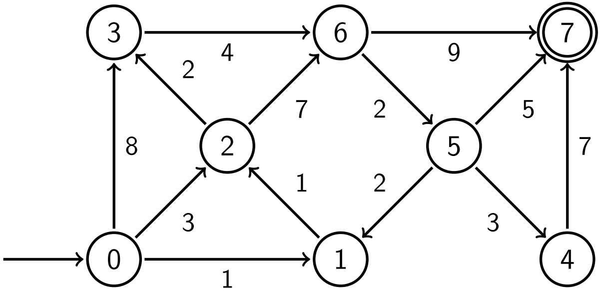
- Task
  - Find the shortest path (*number of hops*)!

- The *simplest* shortest path problem

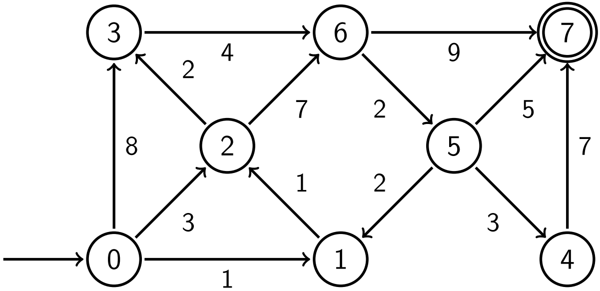


- Task
  - Find the shortest path (*number of hops*)!

- The *classical, non-stochastic, deterministic* shortest path problem

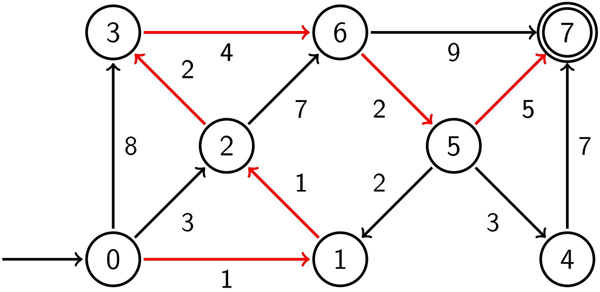


- The *classical, non-stochastic, deterministic* shortest path problem



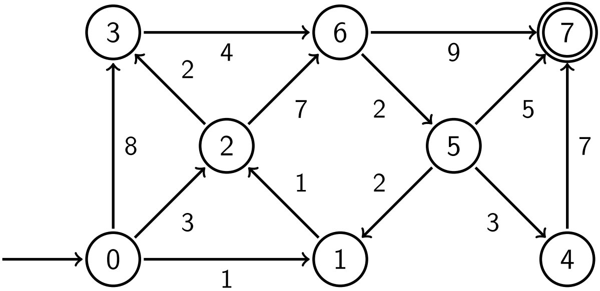
- Task
  - Find the path with the minimal weight sum!

- The *classical*, *non-stochastic*, *deterministic* shortest path problem



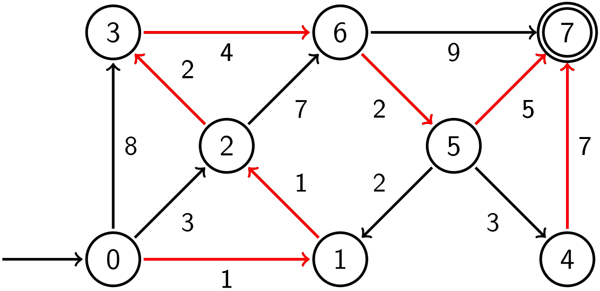
- Task
  - Find the path with the minimal weight sum!

- The *classical, non-stochastic, deterministic* shortest path problem



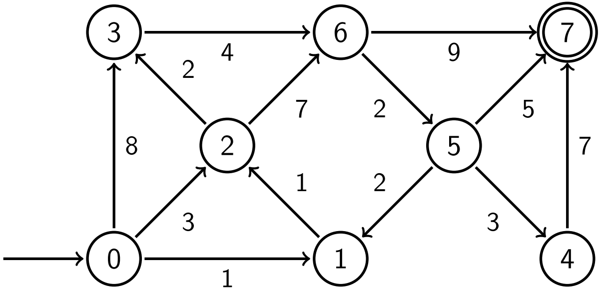
- Task
  - Find the path with the minimal weight sum!
  - Give a strategy to always reach the goal while collecting minimal weight!

- The *classical*, *non-stochastic*, *deterministic* shortest path problem

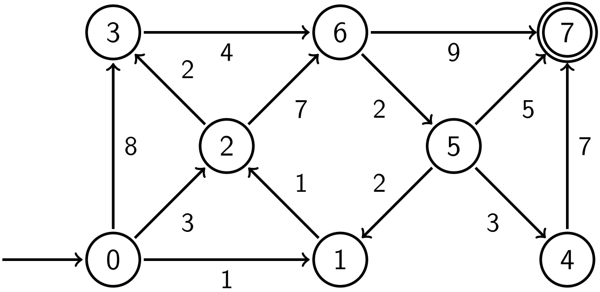


- Task
  - Find the path with the minimal weight sum!
  - Give a strategy to always reach the goal while collecting minimal weight!

- The *stochastic* shortest path problem

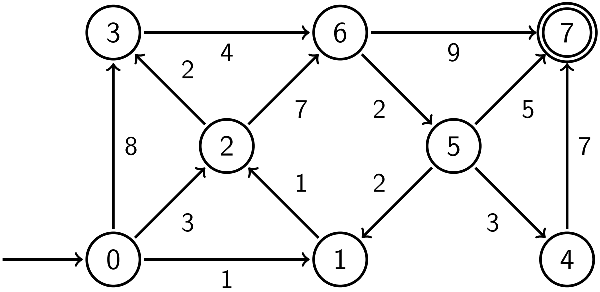


- The *stochastic* shortest path problem



- Markov Decision Process (MDP)

- The *stochastic* shortest path problem



- Markov Decision Process (MDP)
- Task
  - Give a strategy to reach the goal with minimal *expected* accumulated weights!

Section 2

Essential Definitions

- MDP
- Expectation
- Conditional Expectation
- Variance-penalized Expectation
- schedulers, kind of schedulers...

The classic stochastic shortest path problem

- given:
  - a single goal state
  - positive cycle condition: There is no cycle with $\sum wgt \geq 0$
  - goal is reachable from each state
- goal: Maximize the expected accumulated weight until reaching goal state.
- Well known for a long time:
  - There exists an optimal memoryless deterministic scheduler $\mathfrak{S}$.
  - $\mathfrak{S}$ is computable by solving a LP
  - iterative algorithm:
    - start at any feasible scheduler
    - iterative improvement
    - stop at an optimal vertex of the LP (corresponding to some MD scheduler)

Section 3

The classic stochastic shortest path problem

The classic stochastic shortest path problem

- Can we do it better?
- YES! - using **spider construction**!
- Assume furthermore:
  - $\mathcal{M}$ is an MDP with arbitrary integer weights
- The following can be solved in polynomial time:
  - Check: $\mathbb{E}^{inf}_{\mathcal{M},s}(\boxplus\mathrm{goal}) > -\infty$?
  - Compute $\mathbb{E}^{inf}_{\mathcal{M},s}$ if it is finite

Spider Construction

- Idea: construct a new MDP $\mathcal{N}$ from the given MDP $\mathcal{M}$
- Pick a 0-BSCC $\mathcal{E}$ of $\mathcal{M}$ and some vertex s_0 in $\mathcal{E}$.
- $\mathcal{M} \mapsto \mathcal{N} := \mathrm{Spider}_{\mathcal{E},s_0}(\mathcal{M})$
- The spider construction is done by applying the following steps:
  1. Remove all actions $(s, s_\alpha) \in \mathcal{E}$
  2. Add actions (s, τ) for all $s \in \mathcal{E} \setminus \{s_0\}$ such that
     - $P_{\mathcal{N}}(s, \tau, s_0) := 1$
     - $wgt_{\mathcal{N}}(s, \tau) := wgt(s, s_0)$
  3. For each $s \in \mathcal{E} \setminus \{s_0\}$ and $\beta \in \mathrm{Act}_{\mathcal{M}}(s) \setminus \{\alpha_s\}$ let us replace (s, β) by (s_0, β) where
     - $P_{\mathcal{N}}(s_0, \beta, u) := P_{\mathcal{M}}(s, \beta, u)$
     - $wgt_{\mathcal{N}}(s_0, \beta) + wgt(s, s_0) = wgt_{\mathcal{M}}(s, \beta)$

Classification of paths

A path $\pi \in \mathrm{InfPaths}(\mathcal{M})$ is called

- pumping $:\Leftrightarrow \liminf_{n\to\infty}(\mathrm{wgt}(\mathrm{pref}(\pi, \mathrm{n}))) = \infty$
- $\begin{array}{c}\text{(positively)}\\ \text{negatively}\end{array}$ weight divergent $:\Leftrightarrow \begin{array}{c}\limsup_{n\to\infty}\\ \liminf_{n\to\infty}\end{array} = \begin{array}{c}\infty\\ -\infty\end{array}$
- gambling $:\Leftrightarrow \pi$ is positively and negatively weight divergent
- bounded from below $:\Leftrightarrow \liminf_{n\to\infty} \mathrm{wgt}(\mathrm{pref}(\pi, n)) \in \mathbb{Z}$

Classification of end components

We distinguish end components by the following types

- pumping ECs: $\exists$ scheduler $\mathfrak{S} : \mathbb{Pr}(\pi \text{ is pumping}) = 1$
- $\binom{\text{positively}}{\text{negatively}}$ weight divergent ECs: $\exists$ scheduler $\mathfrak{S} : \mathbb{Pr}(\pi \text{ is } \binom{\text{positively}}{\text{negatively}} \text{ weight divergent }) = 1$
- gambling ECs: $\mathbb{E}(\mathrm{MP}) = 0$ and it is positively and negatively weight divergent
- bounded EC: There exists an upper bound and a lower bound

Section 4

Different variants of the stochastic shortest path problem

Imagine to not reach goal with probability 1.

- maximal conditional expected auccumulated reward
- partial expected accumulated reward.

conditional expectation	partial expectation
$\mathbb{CE}$	$\mathbb{PE}$
$\mathbb{CE} = \mathbb{E}(\boxplus \mathrm{goal} \mid \Diamond \mathrm{goal})$	$\pi \not\models \Diamond \mathrm{goal} \Rightarrow \mathrm{wgt}(\pi) := 0$
good approximation for maximizing probability and reward until goal	may lead to quite high $\mathbb{CE}$ paired with low probability of reaching goal

conditional expected accumulated reward

Given:

- MDP $\mathcal{M}$ with non-negative integer weights
- two sets of states $F, G \subseteq \mathrm{States}(\mathcal{M})$

Definition

$$\mathbb{CE}^{max} := \sup_{\mathfrak{S} \in S} (\mathcal{E}^{\mathfrak{S}}_{\mathcal{M}, s_init}(\boxplus F \mid \Diamond G))$$

where S is the set of schedulers:

$$S := \{\mathfrak{S} \mid \mathbb{Pr}^{\mathfrak{S}}_{\mathcal{M}, s_init}(\Diamond G) > 0 \wedge \mathbb{Pr}^{\mathfrak{S}}_{\mathcal{M}, s_init}(\Diamond F \mid \Diamond G) = 1\}$$

results about conditional expected accumulated rewards

- There is a **PTime algorithm** to decide: `Is` $\mathbb{CE}^{max}$ `finite?`
- There is a **pseudo-PTime algorithm** `to calculate an upperbound` $\mathbb{CE}^{ub} \geq \mathbb{CE}^{max}$
- *If we have* $F = G$ *and* $\forall s \in \mathrm{States}(\mathcal{M}) : s \models \exists \Diamond G \Rightarrow \mathrm{Pr}^{min}_{\mathcal{M},s}(\Diamond G) > 0$ there is a **PTime algorithm** `to calculate an upperbound` $\mathbb{CE}^{ub} \geq \mathbb{CE}^{max}$
- The problem `Decide if` $\mathbb{CE}^{max} \bowtie t$ where we have
  - $t \in \mathbb{Q} \ldots$ some rational threshold
  - $\bowtie \in \{<\leq, \geq, >\}$

  is **PSPACE-hard**, solvable in **ExpTime** and *for acyclic MDPs* **PSPACE-complete**
- In **ExpTime** we can compute $\mathbb{CE}^{max}$ `together with an optimal scheduler`

Section 5

Keep an eye on the variance