Jia Ding Sin
Supervision: Tim Quatmann and Darion Haase

WS 2023

1 Introduction

This paper is built on the scientific paper ”The Complexity of Graph-Based Reductions for Reachability in Markov Decision Processes” from Stéphane Le Roux and Guillermo A. Pérez. [1]

2 Preliminaries

2.1 Stochastic Models

Definition 2.1 (Markov chains) *A Markov Chain $\mathcal{C}$ is a tuple (Q,δ) where Q is a finite set of states and $\delta : Q \to \mathbb{D}(Q)$ is a probabilistic transition function, where $\mathbb{D}(Q)$ is the set of all probabilistic distribution on a finite set Q. A probabilistic distribution on Q, is a function $f : Q \to \mathbb{Q}_{\geq 0}$ such that $\sum_{q\in Q} f(q) = 1$.*

A non-empty finite word $\varrho = p_1 p_2 ... p_n$ over Q is defined as a run of Markov Chain. The probability of the run is $\prod_{0\leq i<n}\delta(p_i,p_{i+1})$. ϱ reaches q, if $q = p_i$ for some $0 \leq i \leq n$.

The probability of eventually reaching a set of state $T \subseteq Q$ in $\mathcal{C}$ starting from q_0 can be denoted as $\mathbb{P}_{\mathcal{C}}^{q_0}[\Diamond T]$. $\Diamond T$ is equivalent to $Q \bigcup T$, Q until T. Since not necessary every state in Q is reached before T, a set of states $S \subseteq Q$ is defined, where the path until T only consists of these states. The probability $\mathbb{P}_{\mathcal{C}}^{q_0}[S \bigcup T]$. If $q_0 \in T$, the probability is 1. Otherwise, the probability is calculated as follows:

$$\sum\left\{\prod_{0\leq i<n}\delta(q_i,q_{i+1}) \mid q_0 ... q_{n-1} \in (S \setminus T) \ \&\ q_n \in T \quad \text{for } n \geq 1\right\}.$$

If there exists a set $U \subseteq Q$, where all runs from q_0 to T reaches, the probability of $\mathbb{P}_{\mathcal{C}}^{q_0}[\Diamond T]$ can be reduced to the following lemma:

Lemma 1 *Consider a Markov Chain $\mathcal{C} = (Q,\delta)$ set of states $U, T \subseteq Q$, and a state $q_0 \in Q\backslash U$. If $\mathbb{P}_{\mathcal{C}}^{q_0}[(Q\backslash U)\bigcup T] = 0$, then*

$$\mathbb{P}_{\mathcal{C}}^{q_0}[\Diamond T] = \sum_{u\in U}\mathbb{P}_{\mathcal{C}}^{q_0}[(Q\backslash U)\bigcup u]\mathbb{P}_{\mathcal{C}}^{q_0}[\Diamond T]$$

Definition 2.2 (Markov decision processes) *A (finite and discrete-time) Markov decision processes, MDP, $\mathcal{M}$, is a tuple (Q, A, δ, T) where Q is a finite set of states, A a finite set of actions, $\delta : Q \times A \to \mathbb{D}(Q)$ a probabilistic transition function, and $T \subseteq Q$ a set of target state.*

The notation for the probability of the state p reaching q with the action a, $\delta(p,a)(q)$ will be changed to $\delta(q|p,a)$ for convenience.

Definition 2.3 (Strategies) *A (memoryless deterministic) strategy σ in an MDP $M = (Q, A, \delta, T)$ is a function $\sigma : Q \to A$.*

From MDPs to Chains An MDP $\mathcal{M} = (Q, A, \delta, T)$ and a Markov Chain $M^{\sigma} = (Q, \mu)$, where a strategy σ is applied on. μ is defined as follows: $\mu(q) = \delta(q, \sigma(q))$ for all $q \in Q$. The probability of q, $\mu(q)$, is the same as $\delta(q, \sigma(q))$, where $\sigma(q)$ gives the action assigned to q.

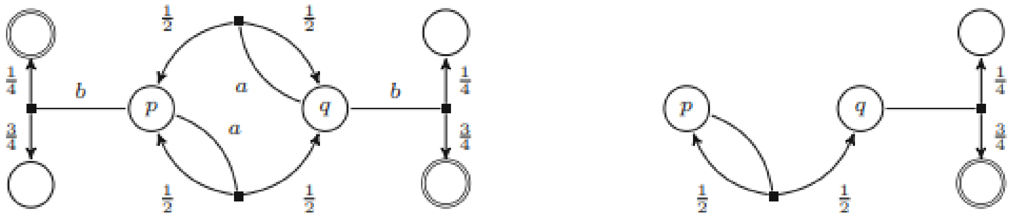

Figure 1: MDP on the left and on the right a Markov chain induced by the left MDP. [1]

The MDP on the left in Figure 1 has the actions $\{a, b\}$. The states are shown as circles, the states with double circles are target states, the squares are distributions. The arrows from states to distributions shows which action is used and the arrows from distributions to states are probabilities.

The strategy σ for the right Markov chain maps as follows: $\{p \mapsto a, q \mapsto b\}$ i.e. $\sigma(p) = a, \sigma(q) = b$. Since the strategy σ is applied on the Markov chain, the arrows with actions are removed. With the Markov chain, the easier probability to calculate would be $\mathbb{P}^{q}_{\mathcal{M}^{\sigma}}[\Diamond T] = \frac{3}{4}$. The probability is written on the arrow from the distributions, after the state q, to the target state.

2.2 Reachability Games Against Nature

It is often difficult to determine the probability $\mathbb{P}^{q_0}_{\mathcal{C}}[\Diamond T]$. That is why families of MDP with the same support, where the probabilities are abstracted, will be considered. The focus is directed to the underlying directed graph from the MDP, where a game against nature is played on. Given an MDP $\mathcal{M} = (Q, A, \delta, T)$, the underlying directed graph $\mathcal{G}_{\mathcal{M}} = (V, E)$ with $V := Q \cup (Q \times A)$ and $E := (q, \langle q, a \rangle) \in Q \times (Q \times A) \cup (\langle p, a \rangle, q) \mid \delta(q|p,a) > 0$. To simplify, the vertex set consists of state vertices and state-action vertices. The edges consist of state to state-action edges $(q, \langle q, a \rangle)$ state-action to next state edges $(\langle p, a \rangle, q)$, where $\delta(q|p,a) > 0$. *Nature* controls the vertices $Q \times A$ in $\mathcal{G}_{\mathcal{M}}$. The game is played on the following arena.

Definition 2.4 (Target Arena) *A target arena is a tuple $\mathcal{A} := (V, V_P, E, T)$ with the bipartite directed graph (V_P, V_N, E), and $T \subseteq V_P$. $V_N := V \backslash V_P$ and for all $u \in V_N$, $uE \neq \varnothing$ with uE being the set of successors of u.*

In this game *Nature* controls the vertices V_N and *Protagonist* controls the vertices V_P. In order to make the arena into an MDP, a family of probability distributions μ using the arena $\mathcal{A}$ will be applied to the MDP. Let the MDP $\mathcal{A}_{\mu}$ be defined as follows:

$\mathcal{A}_\mu = (Q, A, \delta, T)$ with $\mu = (\mu_u \in \mathbb{D}(uE))_{u \in V_N}$, $Q = V_P \uplus \{\bot\}$, $A = V_N$. $\delta(q|p,a)$ is $\mu_a(q)$ if $(p,a)(a,q) \in E$ and otherwise 0. For all $p \in V_P \cup \{\bot\}$ and $a \in A$, $\delta(\bot\,|p,a) = 1$, if $(p,a) \notin E$.

The value of the vertex on the target arena $\mathcal{A}$ differentiate between vertices from V_P and V_N. The maximal reachability probability value of both are defined with respect of full-support probability distributions μ. For $v \in V_P$ the value is defined as $Val^\mu(v) \coloneqq max_\sigma \mathbb{P}^v_{\mathcal{A}^\sigma_\mu}[\Diamond T]$. That means the max reachability probability value of a vertex from the Protagonist is the maximal probability of reaching T from v in $\mathcal{A}^\sigma_\mu$, the target arena induced with the strategy σ and the probability distribution μ. For $u \in V_N$ the value is defined as $Val^\mu(u) \coloneqq \sum\{\mu_u(v) Val^\mu(v) \mid v \in uE\}$, with $Val^\mu(v)$ defined above, since the arena is a bipartite directed graph.

3 The Never-Worse Relation

3.1 Reduction using NWR

3.2 Known Special Cases of NWR

4 Graph-based characterization of the NWR

5 Conclusion

References

[1] S. Le Roux and G. A. Pérez. The complexity of graph-based reductions for reachability in markov decision processes. In C. Baier and U. Dal Lago, editors, *Foundations of Software Science and Computation Structures*, pages 367–383, Cham, 2018. Springer International Publishing.