Multi-Scenario Uncertainty in Markov Decision Processes

Dimitri Scheftelowitsch

May 25, 2018

Markov decision processes (MDPs) are a well-known formalism for modeling discrete event systems. Several extensions of the model have been proposed in order to encompass additional model-level uncertainty in MDPs. Here, we introduce a multi-scenario uncertainty model which has been proposed in the author's PhD thesis and possible future research.

For discrete event systems, Markov decision processes turned out to be a versatile and extensible formalism that captures stochastic behavior, decomposition into *states* with different behaviour, and the possibility to influence the system with a discrete set of *actions* [Put94]. A Markov decision process is defined by a state set $\mathcal{S}$, an action set $\mathcal{A}$, a reward vector $\vec{r} \in \mathbb{R}^{|\mathcal{S}|}$, and a transition probability distribution $\Pr\left[\cdot \mid \cdot, \cdot\right] : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}_{\geqslant 0}$ with $\sum_{t \in \mathcal{S}} \Pr\left[t \mid s, a\right] = 1$ for all $s, a \in \mathcal{S} \times \mathcal{A}$. The control problem for a MDP M is to compute a *policy* $\pi\colon \mathcal{S} \to \mathcal{A}$ that optimizes a measure which maps sequences of (random) rewards $R_1^{(\pi)}, R_2^{(\pi)}, \ldots$ to a *value*. For this *value function*, we choose the *expected discounted reward* $v_\gamma^{(\pi)}(M) = \mathrm{E}\left[\sum_{i \in \mathbb{N}} \gamma^i R_i^{(\pi)}\right]$ for $\gamma \in [0, 1)$.

Over the course of time, several researchers proposed extensions to the basic formalism to capture possible model-level variations and uncertainties [WED94, GLD00]. Here, we focus on a further model of uncertainty discussed in [BS18, Sch17] (and independently, in [RS14] in a model checking context). In this setting, a finite number of MDPs with shared state and action spaces exist with the promise that one of the MDPs models the system adequately.

For this model of uncertainty, a multi-objective perspective similar to the one described in [KKST13] has been chosen: A robust optimization problem can be considered in a multi-objective setting by interpreting the possible *scenarios* as individual optimization goals. Depending on the formulation of the robust problem, its goal can be represented as a policy that optimizes the weighted sum of the optimization goals, or as the set of mutually non-dominating policies.

Stochastic multi-scenario problems In the multi-scenario MDP formalism, we interpret the multi-scenario optimization problem in the following way [BS18, Sch17].

We assume that the uncertainty set $\mathcal{M}$ of MDPs is finite, and the optimization problem is to compute a policy that optimizes the weighted sum $\sum_{M \in \mathcal{M}} w_M v_\gamma^{(\pi)}(M)$ for weights $w_M \in \mathbb{R}$. Unfortunately, this problem is non-linear and non-convex in its nature, and it is shown that the corresponding decision problem is already NP-complete for two scenarios.

In order to provide a practically usable solution for the optimization problem, we have considered several approaches which can be roughly grouped into two main directions. The first direction was to formulate the optimization problem as an instance of mathematical programming and to use industrial-grade solvers. Three formulations have been derived: non-linear programming with and without gradients and integer linear programming. The second direction was to use local search algorithms, ideologically similar to policy iteration, in order to compute a locally optimal policy.

As all these approaches describe heuristic algorithms, an empirical comparison of the formulations has been undertaken, resulting in several observations. The first observation that could have been made was that, unsurprisingly, giving the solver more information about the problem structure (such as the gradient of the objective function or the integer programming formulation), results in less time spent on solving.

The second observation was that the local optimization heuristics offer an acceptable solution quality with a large benefit in solution time. While the gains in solution time were not surprising, the relative difference in the objective value, compared to the results of the exact NLP or ILP solvers, was in most cases under 3%, which is noteworthy, as it allows one to use the local optimization approach in practice without sacrificing much in terms of solution quality.

Future directions The work done in the thesis and the supporting paper can be extended along several directions.

- The local optimization heuristic, which is mainly based on local search, could be extended to (and compared with) simulated annealing. This approach admits a probability of order e^{-n} of selecting a n-th candidate solution even if it is worse than the current one in order to escape from local optima; the comparison to the current heuristic which computes only local optima would be interesting.

- A straightforward extension is to consider the tuple $(v_\gamma^{(\pi)}(M))_{M \in \mathcal{M}}$ as a value in a multi-objective value space and to consider the non-dominated set enumeration problem. We believe it should be possible to derive a policy-iteration-like enumeration algorithm and, furthermore, an enumeration algorithm for the convex hull of the non-dominated solution in value space.

- Another direction is to consider the multi-scenario problem from a robust perspective and solve the problem $\max_\pi \min_{M \in \mathcal{M}} v_\gamma^{(\pi)}(M)$. A similar problem has been solved for a continuous scenario space [WED94, GLD00]; our intuition is that the robust problem will be harder in our setting, as several assumptions which are valid in [WED94, GLD00] (most prominently, the rectangularity assumption) do not hold here.

References

[BS18] Peter Buchholz and Dimitri Scheftelowitsch. Computation of weighted sums of rewards for concurrent MDPs. 2018. In review.

[GLD00] Robert Givan, Sonia M. Leach, and Thomas L. Dean. Bounded-parameter Markov decision processes. *Artif. Intell.*, 122(1-2):71–109, 2000.

[KKST13] Kathrin Klamroth, Elisabeth Köbis, Anita Schöbel, and Christiane Tammer. A unified approach for different concepts of robustness and stochastic programming via non-linear scalarizing functionals. *Optimization*, 62(5):649–671, 2013.

[Put94] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York, NY, USA, 1994.

[RS14] Jean-François Raskin and Ocan Sankur. Multiple-environment Markov decision processes. In *34th International Conference on Foundation of Software Technology and Theoretical Computer Science, FSTTCS 2014, December 15-17, 2014, New Delhi, India*, pages 531–543, 2014.

[Sch17] Dimitri Scheftelowitsch. *Markov Decision Processes With Uncertain Parameters*. PhD thesis, TU Dortmund, TU Dortmund, November 2017. Submitted for review.

[WED94] Chelsea C. White and Hany K. El-Deib. Markov decision processes with imprecise transition probabilities. *Operations Research*, 42(4):739–749, 1994.