Planning Using a Portfolio of Reduced Models

Extended Abstract

Sandhya Saisubramanian, Shlomo Zilberstein, and Prashant Shenoy
College of Information and Computer Sciences
University of Massachusetts, Amherst
{saisubramanian,shlomo,shenoy}@cs.umass.edu

ABSTRACT

Existing reduced model techniques simplify a problem by applying a uniform principle to reduce the number of considered outcomes for all state-action pairs. It is non-trivial to identify which outcome selection principle will work well across all problem instances in a domain. We aim to create reduced models that yield near-optimal solutions, without compromising the run time gains of using a reduced model. First, we introduce *planning using a portfolio of reduced models*, a framework that provides flexibility in the reduced model formulation by using a portfolio of outcome selection principles. Second, we propose *planning using cost adjustment*, a technique that improves the solution quality by accounting for the outcomes ignored in the reduced model. Empirical evaluation of these techniques confirm their effectiveness in several domains.

KEYWORDS

Reasoning in Agent-based Systems; Single and Multi-Agent Planning and Scheduling

ACM Reference Format:
Sandhya Saisubramanian, Shlomo Zilberstein, and Prashant Shenoy. 2018. Planning Using a Portfolio of Reduced Models. In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018,* IFAAMAS, 3 pages.

1 INTRODUCTION AND BACKGROUND

A reduced or simplified model of a large Markov Decision Process (MDP) helps to cope with the complexity of solving large stochastic planning problems [10]. Reduced models simplify the problem by partially or completely ignoring the uncertainty, thereby reducing the set of reachable states a planner needs to consider [3–5, 7, 9, 11, 13–15]. While the existing reduced model techniques accelerate the planning process, they do not guarantee bounded-optimal performance and it is often hard to predict when they will work particularly well. For example, consider a robot navigating through a building. A plan generated by a simple reduced model might work well when the robot is moving through uncluttered region, but a more informative reduced model or the full model may be required to reliably navigate through a narrow corridor [12]. Hence, we consider reduced models with different levels of detail, created by different outcome selection principles.

We introduce two frameworks that help create reduced models that efficiently balance the trade-off between solution quality and planning time. First, we introduce *planning using a portfolio of reduced models*, a framework that provides flexibility in switching between different outcome selection principles to customize reduced models. Second, we introduce *planning using cost adjustment*, a technique that improves the solution quality of reduced models by altering the costs of actions to account for the consequences of ignored outcomes in the reduced model.

Consider a Stochastic Shortest Path (SSP) MDP defined by $M = \langle S, A, T, C, s_0, S_G \rangle$, and let $\theta(s,a)$ denote the set of all outcomes of action a in state s in M, $\theta(s,a) = \{s' | T(s,a,s') > 0\}$.

A **reduced model** of an SSP M is represented by the tuple $M' = \langle S, A, T', C, s_0, S_G \rangle$ and characterized by an altered transition function T' such that $\forall (s,a) \in S \times A, \theta'(s,a) \subseteq \theta(s,a)$, where $\theta'(s,a) = \{s' | T'(s,a,s') > 0\}$ denotes the set of outcomes in the reduced model for action a in state s. That is, the reduced model considers a modified transition function with a subset of the outcomes for each (s,a) pair. In this work, we normalize the probabilities of the outcomes included in the reduced model so that they sum to one. The outcome selection process in a reduced model framework accounts for two decisions: the number of outcomes, and how the outcomes are selected. An *outcome selection principle* (OSP) performs the outcome selection process per (s,a) pair in the reduced model, thus determining the transition function for the (s,a) pair. The OSP can be some simple function such as always choosing the most likely outcome or a more complex function. Traditionally, a reduced model is characterized by a single OSP. That is, a single principle is used to determine the number of outcomes and how the outcomes are selected across the entire model.

2 PORTFOLIO OF REDUCED MODELS

We define a generalized framework, *planning using a portfolio of reduced models*, that facilitates the creation of reduced models that can better capture the domain features by switching between different OSPs, each of which represents a different reduced model. The framework is inspired by the benefits of using portfolios of algorithms to solve complex problems [8].

Definition 2.1. Given a portfolio of finite outcome selection principles, $Z = \{\rho_1, \rho_2, ..., \rho_k\}$, k>1, a **model selector**, Φ, generates T' for a reduced model by mapping every state-action pair to an OSP, $\Phi : S \times A \rightarrow \rho_i, \rho_i \in Z$, such that $T'(s,a,s') = T_{\Phi(s,a)}(s,a,s')$, where $T_{\Phi(s,a)}(s,a,s')$ denotes the transition probability corresponding to the OSP selected by the model selector.

Trivially, model selectors used by the existing reduced models are a special case of the above definition, as Φ always selects the same ρ_i for every (s,a) pair. Typically, in *planning using a portfolio of reduced models* (PRM), the model selector utilizes more than one

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.

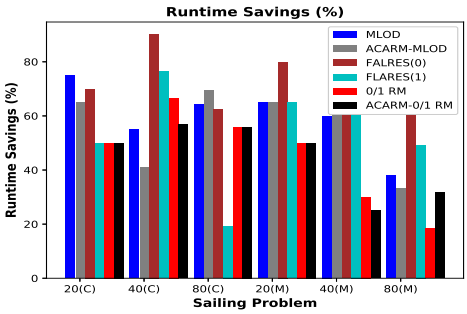

Figure 1: Sailing Domain – Runtime Savings (%)

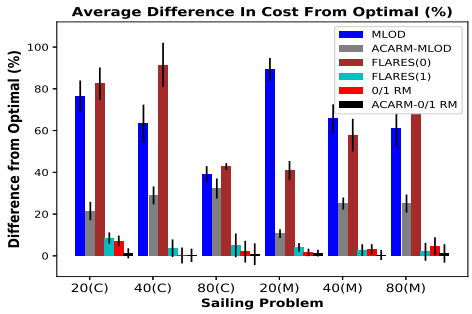

Figure 2: Sailing Domain – Cost Difference (%)

outcome selection principle to determine T'. Each (s, a) pair may have a different number of outcomes and a different mechanism to select the specific outcomes. For example, for a certain (s, a) pair, the model selector may select the most likely outcome, and for another (s, a) pair, it may greedily select two outcomes based on the heuristic values. Although the model selector could use multiple ρ_i, the resulting reduced model is still an SSP.

A **0/1 reduced model** (0/1 RM) is a PRM with a model selector that selects either one or all the outcomes of an action in a state to be included in the reduced model. A 0/1 RM is characterized by a model selector, $\Phi_{0/1}$, that either ignores the stochasticity completely (0) by considering only one outcome of (s, a), or fully accounts for the stochasticity (1) by considering all the outcomes of (s, a) in the reduced model. Thus, for every SSP, there exists a 0/1 reduced model that guarantees goal reachability with probability 1, if a proper policy exists in M. In the worst case, devising an efficient Φ for a PRM may involve evaluating every $\rho_i \in Z$. Let τ_{max} denote the maximum time taken for this evaluation. Then the worst case time complexity for a Φ to generate T' for a PRM and 0/1 RM are $O(|A| \cdot 2^{|S|} \cdot \tau_{max})$ and $O(|A| \cdot |S|^2 \cdot \tau_{max})$, respectively. These bounds underscore the need for developing faster techniques to evaluate and identify relevant OSPs. Approximate model selectors that reasonably balance the trade-off can be computed using heuristics or based on state features.

3 COST ADJUSTMENT

One of the reasons for the sub-optimality of existing reduced model approaches is that certain outcomes are completely ignored. We introduce *planning using cost adjustment*, a technique that accounts for the ignored outcomes by adjusting the action costs in the reduced model, thus resulting in near-optimal action selection.

Definition 3.1. A **cost adjusted reduced model** (CARM) of an SSP M is a reduced model represented by the tuple $M'_{ca} = \langle S, A, T', C', s_0, S_G \rangle$ and characterized by an altered cost function C' such that $\forall (s, a)$ in reduced model,

$$C'(s, a) \leftarrow Q^*(s, a) - \sum_{s' \in \theta'(s,a)} T'(s, a, s')V^*(s').$$

In the above definition, $Q^*(s, a)$ and $V^*(s')$ denote the optimal Q-value and optimal value of the action in M. Given an SSP and its reduced model (not necessarily PRM), the costs are adjusted for every (s, a) in the reduced model to account for the ignored outcomes. Since the costs are adjusted based on the difference in values of states, this may lead to negative cost cycles in the reduced model. Therefore, the necessary and sufficient condition for non-negative cost in a CARM is that the transition function in the reduced model satisfies $Q^*(s, a) \geq \sum_{s' \in \theta'(s,a)} T'(s, a, s')V^*(s')$.

Proposition 3.2. *A CARM that preserves goal reachability yields optimal action selection for the SSP.*

Planning with Approximate Cost Adjustment. Since generating a CARM may involve solving the SSP to estimate the optimal values of the outcomes, we propose an approximation that estimates the costs in the reduced model. A **feature-based cost function** estimates the cost of an action in a state using the features of the state $\vec{f}(s)$, $C'(s, a) = g(\vec{f}(s), a)$, with $g : \vec{f} \times A \rightarrow \mathcal{R}$. The resultant reduced model with approximate costs is referred to as *approximately cost adjusted reduced model* (ACARM). Given a set of state features that significantly affect the cost of actions, the feature-based costs are estimated by generating and solving sample problems. The cost adjustments are computed for the samples using their exact solutions and the feature-based costs are learned and projected onto the target problem.

4 RESULTS AND CONCLUSION

We present the results (Figures 1 and 2) of our approach on the sailing domain [6]. All results are averaged over 100 trials and the average times include the time spent on re-planning. ACARM-MLOD and ACARM-0/1 RM use feature-based costs. The deterministic problems are solved using the A* algorithm [2], and the others are solved using LAO* [1], and complemented by re-planning when necessary. The lower difference in cost values indicates that the performance of the technique is near optimal. In most problems, ACARM-0/1 RM yields almost optimal results without compromising the run time gains of using a reduced model. We have obtained similar results in several other domains.

In summary, we propose two general methods that help create robust reduced models of large SSPs. Our results contribute to a better understanding of how disparate reduced model techniques could be used together to leverage their complementary benefits.

ACKNOWLEDGMENTS

Support for this work was provided in part by the U.S. National Science Foundation grant No. 1524797.

REFERENCES

[1] Eric A. Hansen and Shlomo Zilberstein. 2001. LAO*: A heuristic search algorithm that finds solutions with loops. *Artificial Intelligence* 129 (2001), 35–62.

[2] Peter E. Hart, Nils J. Nilsson, and Bertram Raphael. 1968. A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Transactions on Systems Science and Cybernetics* 4 (1968), 100–107.

[3] Jörg Hoffmann. 2001. FF: The Fast-Forward Planning System. *AI Magazine* 22 (2001), 57–62.

[4] Thomas Keller and Patrick Eyerich. 2011. A Polynomial All Outcome Determinization for Probabilistic Planning. In *Proceedings of the 21st International Conference on Automated Planning and Scheduling*. 331–334.

[5] Emil Keyder and Hector Geffner. 2008. The HMDP Planner for Planning with Probabilities. In *Proceedings of the International Planning Competition (IPC 2008)*.

[6] Levente Kocsis and Csaba Szepesvári. 2006. Bandit based Monte-Carlo planning. In *Proceedings of the 17th European Conference on Machine Learning*. 282–293.

[7] Andrey Kolobov and Daniel S. Weld. 2009. ReTrASE: Integrating Paradigms for Approximate Probabilistic Planning. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence*. 1746–1753.

[8] Marek Petrik and Shlomo Zilberstein. 2006. Learning Parallel Portfolios of Algorithms. *Annals of Mathematics and Artificial Intelligence* 48, 1-2 (2006), 85–106.

[9] Luis Pineda, Yi Lu, Shlomo Zilberstein, and Claudia V. Goldman. 2013. Fault-tolerant Planning under Uncertainty. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*. 2350–2356.

[10] Luis Pineda and Shlomo Zilberstein. 2014. Planning Under Uncertainty Using Reduced Models: Revisiting Determinization. In *Proceedings of the 24th International Conference on Automated Planning and Scheduling*. 217–225.

[11] Sandhya Saisubramanian, Shlomo Zilberstein, and Prashant Shenoy. 2017. Optimizing Electric Vehicle Charging Through Determinization. In *Proceedings of the Scheduling and Planning Applications woRKshop (SPARK), 27th International Conference on Automated Planning and Scheduling*. 15–23.

[12] Breelyn Styler and Reid Simmons. 2017. Plan-Time Multi-Model Switching for Motion Planning. In *Proceedings of the 27th International Conference on Automated Planning and Scheduling*. 558–566.

[13] Sungwook Yoon, Alan Fern, and Robert Givan. 2007. FF-Replan: A Baseline for Probabilistic Planning. In *Proceedings of the 17th International Conference on Automated Planning and Scheduling*. 352–359.

[14] Sungwook Yoon, Alan Fern, Robert Givan, and Subbarao Kambhampati. 2008. Probabilistic Planning via Determinization in Hindsight. In *Proceedings of the 23rd Conference on Artificial Intelligence*. 1010–1016.

[15] Sungwook Yoon, Wheeler Ruml, J. Benton, and Minh B. Do. 2010. Improving determinization in hindsight for on-line probabilistic planning. In *Proceedings of the 20th International Conference on Automated Planning and Scheduling*. 209–216.