Recent Advances in Symbolic Dynamic Programming for Hybrid MDPs and POMDPs

Scott Sanner & Zahra Zamani
NICTA & ANU
Canberra, Australia
{scott.sanner,zahra.zamani}@nicta.com.au

May 10, 2013

Abstract

Many real-world decision-theoretic planning problems are naturally modeled using mixed discrete and continuous state, action, and observation spaces, yet little work has provided *exact* methods for performing exact dynamic programming backups in such problems. This overview talk will survey a number of recent developments in the exact and approximate solution of mixed discrete and continuous (hybrid) MDPs and POMDPs via the technique of symbolic dynamic programming (SDP).

1 Overview

The core idea between symbolic dynamic programming (SDP) is simple: express the components of an MDP or a POMDP in some form for which all operations required in a dynamic programming solution are closed-form. When examining MDPs and POMDPs with continuous components (especially state, but also actions and observations), this becomes difficult for the primary reason that the required integrals cannot always be computed in closed-form. Furthermore, even if these integrals could be solved in a closed form, the overall value function representation computed in dynamic programming often does not remain tractably sized for naïve representations. Hence, effective SDP requires not only a functional form which all operations can be computed symbolically and in closed-form, but it also typically requires a data structure to help maintain a compact form and support efficient dynamic programming operations.

A number of advances have been made in recent years regarding the application of SDP ideas to mixed discrete and continuous (hybrid) MDPs and POMDPs. This work aims to survey these advances for an audience familiar with MDPs and POMDPs with the aim of providing a new set of tools to researchers for either the direct solution of MDPs and POMDPs or components to model and perform required operations in model-based reinforcement learning.

In brief, the overview begins with [1], which provides a closed-form exact value function and policy for discrete action, discrete noise, continuous state piecewise nonlinear transition and reward MDPs and also introduces the XADD data structure that helps maintain a compact representation of the value function. This marks the first time that optimal value functions and policies could be derived for all states for *general piecewise nonlinear* transition systems and reward. One next step was to extend this work to continuous actions as done in [3], however for the more restricted case of continuous action, discrete noise, continuous state piecewise linear transition and reward MDPs. An alternate next step was to extend these results to the case of continuous observations in POMDPs as provided in [4]. A crucial recent contribution was to relax the restriction on discrete noise to allow general continuous noise (e.g., state-dependent Gaussian noise), which was done in [2] via a robust solution. Together, these contributions cover important recent advances in the dynamic programming solution of hybrid MDPs and POMDPs.

References

[1] S. Sanner, K. V. Delgado, and L. Nunes de Barros. Symbolic dynamic programming for discrete and continuous state MDPs. In *In Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence (UAI-11)*, Barcelona, Spain, 2011.

[2] Z. Zamani, S. Sanner, K. V. Delgado, and L. Nunes de Barros. Robust optimization for hybrid mdps with state-dependent noise. In *In Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI-13)*, Beijing, China, 2013.

[3] Z. Zamani, S. Sanner, and C. Fang. Symbolic dynamic programming for continuous state and action mdps. In *In Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI-12)*, Toronto, Canada, 2012.

[4] Z. Zamani, S. Sanner, P. Poupart, and K. Kersting. Symbolic dynamic programming for continuous state and observation pomdps. In *In Proceedings of the 26th Annual Conference on Advances in Neural Information Processing Systems (NIPS-12)*, Lake Tahoe, Nevada, 2012.