Representation Discovery for MDPs Using Bisimulation Metrics

Sherry Ruan, Gheorghe Comanici, Prakash Panangaden, Doina Precup
School of Computer Science, McGill University, Montreal

Introduction

Markov Decision Processes (MDPs) are a powerful mathematical model widely adopted in planning and learning under uncertainty. Solving large sequential decision problems modelled as MDPs requires the use of approximations to represent the state space. One approach is inspired by work on **bisimulation relations** [Larsen and Skou 1989] and their relaxation, **bisimulation metrics** [Desharnais, et al. 1999]. Our contribution is based on a novel, flexible, iterative refinement algorithm to automatically construct approximate state space representations for MDPs. Moreover, the framework provides susbstatial computational improvements for bisimulation metrics and relations. Theoretical results guarantee convergence to desired relations/metrics, and empirical work on simulated models illustrate the accuracy and savings (in time and memory usage) of the new approach.

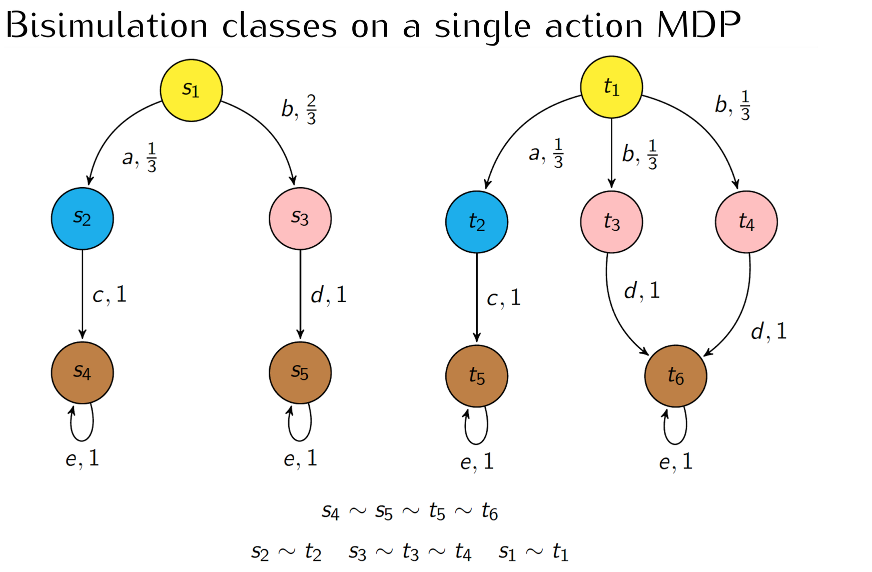
Background

Probabilistic Bisimulation

Given a Markov Decision Process (S, A, R, P, γ), a **probabilistic bisimulation relation** is an equivalence relation $\sim$ on the state space S such that for all equivalent pairs $s \sim t$, all equivalence classes $C \in S/\sim$, and actions $a \in A$,

(1) $R^a(s) = R^a(t)$ (2) $P^a(s)(C) = P^a(t)(C)$

Bisimulation classes on a single action MDP



Bisimulation Metrics

- **Bisimulation metrics** are a quantitative analogue of **bisimulation relations**
- Metrics are smooth with respect to changes on the transition probabilities over the state space
- The conditions (1) and (2) above are replaced by the following numerical condition: *A distance function d over S is **a bisimulation metric** if, for every pair s, t, $s \sim t \iff d(s,t) = 0$.*

Kantorovich metrics for computing bisimulation

- One could use the **Kantorovich metric** $\mathcal{T}(d,\cdot,\cdot)$ to find the *optimal* coupling of two probability distributions (for a numerical relaxation to condition (2))
- Based on the Kantorovich-Rubinstein Duality Theorem, we adopt **Earth Mover's Distance (EMD)** algorithms [Rubner, Tomasi, and Guibas 2000] to compute a transportation problem equivalent to the Kantorovich distance $\mathcal{T}(d,\cdot,\cdot)$.

Let F be an operator on the set of all metrics on S, defined as

$$F(d) : (s,t) \mapsto \max_a \left(|R^a(s) - R^a(t)| + \gamma \mathcal{T}(d, P^a(s), P^a(t)) \right)$$

If d is a fixed point of F then d is a **bisimualtion metric**.

Bisimulation metrics are attractive because they allow quantifying approximation errors for *any* state space representation. **However, metric computation is very expensive!** [Ferns et al. 2014] [Ferns and Precup 2014]

Theoretical Contriubtion

We characterize the same bisimulation metric in three equivalent ways:

Fixed point characterization: $d^* = \sup_n F^n(0)$ (where F is defined as above)

We say a matrix B is a partition if all entries in B are in $\{0,1\}$ and $B^T\mathbf{1} = \mathbf{1}$. Let $\{B_n\}_{n=1}^{\infty}$ be a sequence of partitions such that:

- $B_0 = \mathbf{1}$
- $R^a \in colspan(B_n)$, $\forall a \in A$
- $P^a\phi \in colspan(B_n)$, $\forall a \in A, \forall \phi \in colspan(B_{n-1})$

Given this sequence, choose s and s' with $\phi(s) = 1$ and $\phi(s') = 1$ and define

$$d_n : (\phi, \phi') \mapsto \max_a \left(\left|R^a(s) - R^a(s')\right| + \gamma \mathcal{T}(d_{n-1}, B_{n-1}P^a(s), B_{n-1}P^a(s')) \right)$$

Supremum over metrics on partitions: $d^* = \sup_n B_n d_n B_n^T$.

Let $\{B_n, C_n\}_{n=1}^{\infty}$ be a sequence of partitions paired with subsets $C_n \subset colset(B_{n-1})$. Let Π_n be the orthogonal projection on C_n and perform the construction using as conditions

- $\Pi_n R^a \in colspan(B_n)$, $\forall a \in A$
- $\Pi_n P^a \phi \in colspan(B_n)$, $\forall a \in A, \forall \phi \in colspan(B_{n-1})$
- $colspan(B_{n-1}) \subset colspan(B_n)$

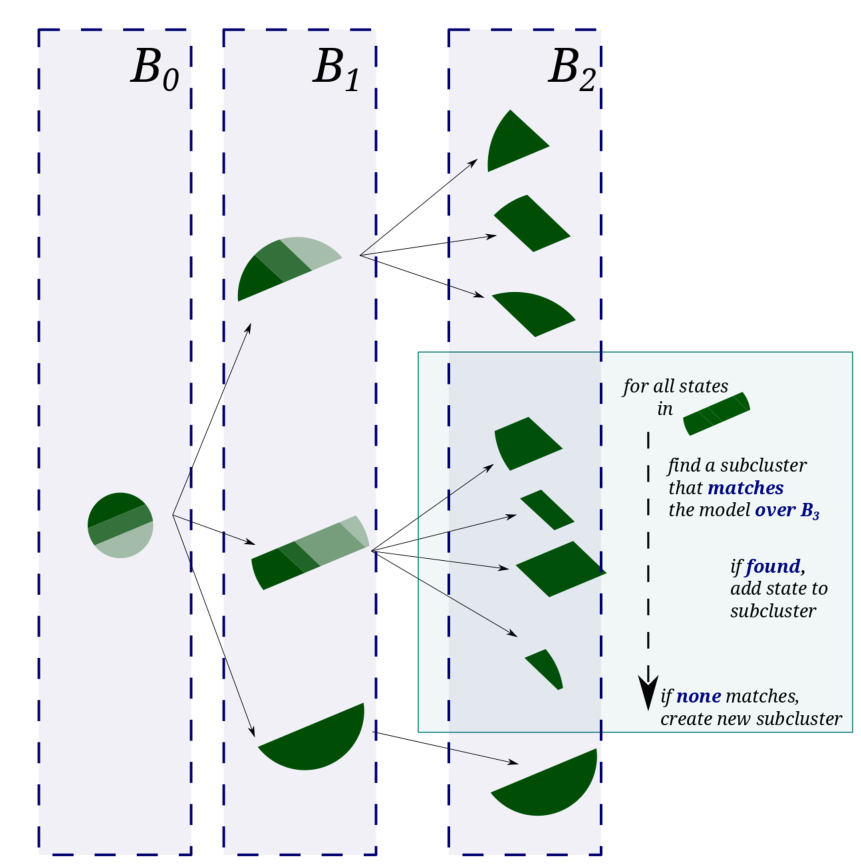
This defines bisimulation metrics as a **supremum over metrics on asynchronous partitions**.

Algorithmic Contribution

The additional characterizations provide alternative ways of computing approximations to the bisimulation metric d^*. We do this using an iterative algorithm which generates partition refinements B_n and computes a metric over these partitions. An efficient refinement of B_n into B_{n+1} is described to the right. Note that the proposed refinement is local (dependent on a particular cluster we wish to refine). The final complexity of computing d_n given d_{n-1} depends on the following terms:

- $\phi^T\phi$, the size of the block ϕ
- B_ϕ, the set subclusters of ϕ
- B_n, the size of the partition after n iterations

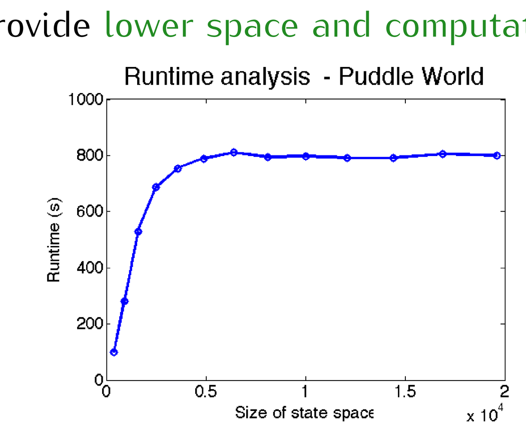
$$O\left(\sum_{\phi \in B_{n-1}} (\phi^T\phi)|B_\phi||A|\right) + O\left(|B_n|^2|B_{n-1}|^2 \log|B_{n-1}||A|\right)$$



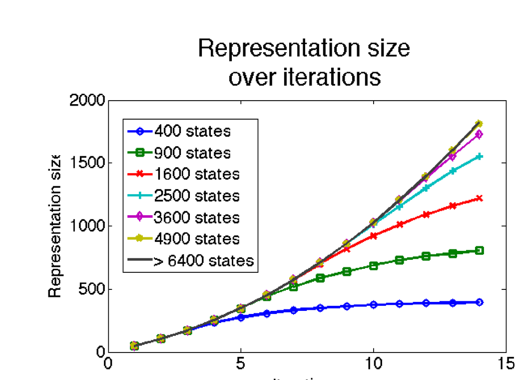
Empirical Results

Experimental results illustrate that partial metric computations

- Maintain strong convergence properties
- Guide the representation search under flexible heuristic choice
- Provide lower space and computational complexity



A plot of the runtime as a function of state space size when computing the metric. If metrics are computed over the state space instead, the runtime jumps from 129 seconds on a 400 states environment, to 1375 seconds when the number of states is 1600.

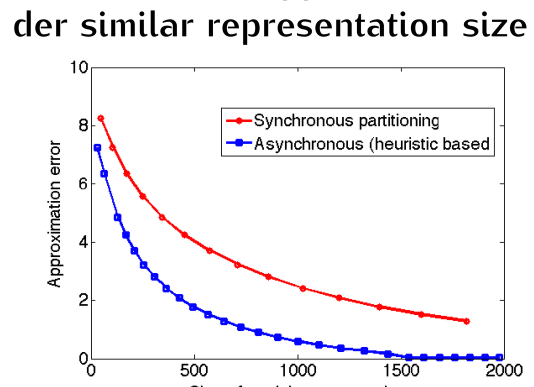


The number of features in the intermediate steps of the algorithm. Note that for state spaces larger than 4900, the number of features does not change substantially with the size of the space.

Asynchronous computation: A plot of the approximation error in the value function computation (L_∞ norm) as the size of the alternative representation increases. This plot was generated **on a Puddle World of size 4900**.

Asynchronous heuristics: For each partition we used a **heuristic** which selects the largest block first to update the partition and metric. As can be seen in Figure 4, the asynchronous algorithm obtains representations of higher quality in much earlier stages of the iterative framework.

Value function approximation under similar representation size



Conclusion

- **New theoretical characterizations** for bisimualtion relations and metrics
- **A novel flexible iterative refinement algorithm** to automatically construct an approximate state space representation for Markov Decision Processes
- We addressed the main drawback for previous approaches: *the expensive computation the bisimulation metrics*

References

Ferns, N., and Precup, D. 2014. Bisimulation metrics are optimal value functions. UAI.
Desharnais, J.; Gupta, V.; Jagadeesan, R.; Panangaden, P. 1999. Metrics for labeled Markov systems. CONCUR
Larsen, K. G.; Skou, A. 1989. Bisimulation through probabilistic testing. ACM SIGPLAN-SIGACT
Rubner, Y.; Tomasi, C.; and Guibas L. J. 2000. The earth movers distance as a metric for image retrieval IJCV.