Representation Discovery for MDPs Using Bisimulation Metrics

Sherry Shanshan Ruan
Supervisor: Professor Prakash Panangaden
Reasoning and Learning Lab, School of Computer Science, McGill University

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 Markov Decision Processes (MDPs) requires the use of approximations to represent the state space. One approach for this problem is to use bisimulation relations [Larsen and Skou 1989] or their relaxation as bisimulation metrics [Desharnais, Gupta, Jagadeesan, and Panangaden 1999]. However, bisimulation metric computation is very expensive. In this work, we tackle this problem by **proposing a significant improvement in how bisimulation metrics are computed**.

Background

Probabilistic Bisimulation

Given a Markov Decision Process (S, A, R, P, γ), a **probabilistic bisimulation relation** is an equivalence relation $\sim$ on S such that if $s \sim t$ then

$$(1)\ \forall a \in A, R_{sa} = R_{ta} \qquad (2)\ \forall a \in A, \forall c \in S/\sim, P^a_{sc} = P^a_{tc}$$

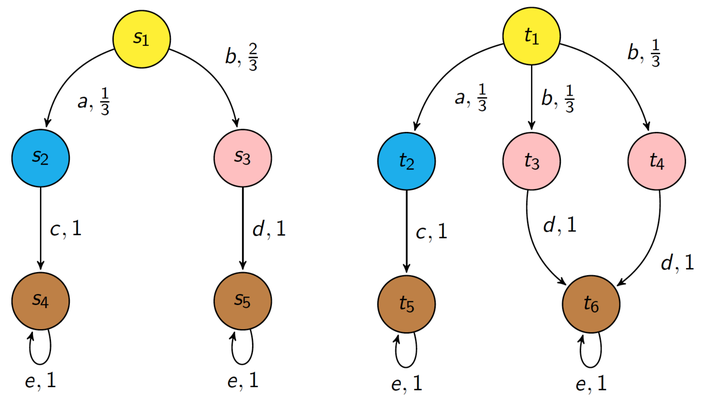

Fig. 1: An example of probabilistic bisimulation relations

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
- We use the **Kantorovich metric** to find the **optimal** coupling of two probability distributions
- **Kantorovich metric** can be transformed into transportation problems by Kantorovich-Rubinstein Duality (max $\rightarrow$ min)
- We adopt **Earth Mover's Distance (EMD)** algorithms [Rubner, Tomasi, and Guibas 2000] to find the **minimal** cost for transforming one histogram into the other

Let s and s' be two states, a **bisimulation metric** is a fixed point of of F defined as follows

$$F(d)(s, s') := \max_a \left(|R_{sa} - R_{s'a}| + \gamma \mathcal{T}(d, P_{sa}, P_{s'a})\right)$$

However, bisimulation metric computation is very expensive! [Ferns and Precup 2014]

Theoretical and Algorithmic Contriubtion

We characterize the same bisimulation metric in three equivalent ways:

- Supremum over $F^n(0)$ (where F is defined above)
- Supremum over metrics on partitions
- Supremum over asynchronous declustering partitions.

Algorithm 1 Partition declustering

Given a partition $B_n, n \geq 1$
$B_{n+1} \leftarrow \emptyset$
for all $\phi \in B_n$ **do**
  $B_\phi \leftarrow \emptyset$
  for all s with $\phi(s) = 1$ **do**
    for all $\phi' \in B_\phi$ **do**
      choose s' with $\phi'(s') = 1$
      if $\forall a, \forall \phi'' \in B_{n-1}, (\mathcal{P}^a\phi'')(s) = (\mathcal{P}^a\phi'')(s')$
      then $\phi'(s) \leftarrow 1$
    end for
    if $\phi'(s) = 0,\ \forall \phi' \in B_\phi$
    then add a new element $\hat{\phi}$ to B_ϕ and set $\hat{\phi}(s) = 1$
  end for
  add the elements of B_ϕ to B_{n+1}
end for

- A partition $\mathcal{B}$ is defined as a basis $\{\phi_i\}_{i=1}^m$ s.t. $\phi_i \in \{0, 1\}$ and $\sum_i \phi_i(s) = 1, \forall$ state $s \in S$.
- $\mathcal{P}^a$ is associated with the policy choosing action a deterministically.

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

As n approaches ∞, the update algorithm runs in $O(|A|(|S||B_\sim| + |B_\sim|^4 \log|B_\sim|))$, which is an upper bound for the update at any step.

Empirical Results

Experimental results illustrate that partial metric computations

- Maintain strong convergence properties
- Guide the representation search
- Provide much lower space and computational complexity

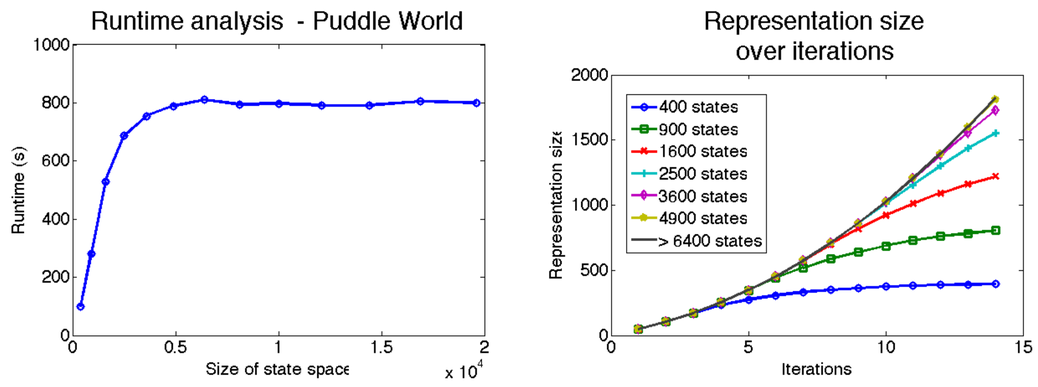

Fig. 3: ***Puddle World - computing the metric.*** **Left:** A plot of the runtime as a function of state space size when computing the metric. If metrics are computed over the state space instead, the runtime jumps from 129 seconds on a 400 states environment, to 1375 seconds when the number of states is 1600. **Right:** 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.

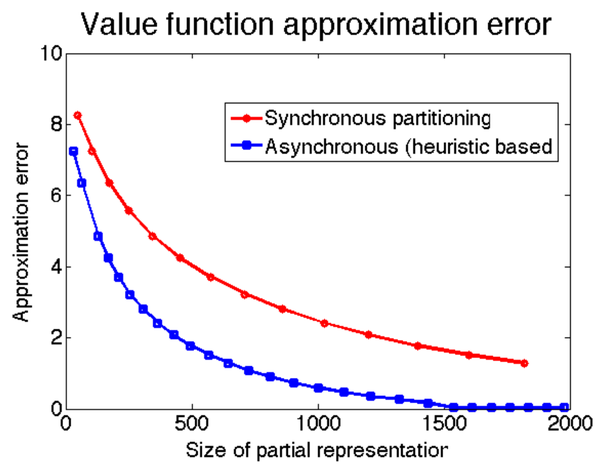

Fig. 4: **Asynchronous computation:** A plot of the approximation error in the value function computation (L_∞ norm) as the size of the alternative representation increases. This plot was generated on a Puddle World of size 4900.

Asynchronous Algorithm: 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.

Conclusion

- We provided a novel flexible iterative refinement algorithm to automatically construct an approximate state space representation for Markov Decision Processes
- We addressed a drawback of the previous approach [Comanici, Panangaden, and Precup 2012] which is the expensive computation of the bisimulation metrics
- We proposed an algorithm to generate an iteratively improving sequence of state space partitions
- We provided experimental illustrations of the accuracy and savings (in time and memory usage) of the new algorithm, compared to traditional bisimulation metric computation

Futher Work

Our approach opens the door to more specialized strategies to finding bisimulation-based MDP representations:

- Extend experiments on synthetic data to real data with large state space
- Investigate more sophisticated heuristics based on different problems

References

[1] Comanici, G.; Panangaden, P.; and Precup, D. 2012. On-the-Fly Algorithms for Bisimulation Metrics. In QEST.
[2] Ferns, N., and Precup, D. 2014. Bisimulation metrics are optimal value functions. In UAI.
[3] Desharnais, J.; Gupta, V.; Jagadeesan, R.; Panangaden, P. 1999. Metrics for labeled Markov systems. In CONCUR'99 Concurrency Theory (pp. 258-273). Springer Berlin Heidelberg.
[4] Larsen, K. G.; Skou, A. 1989. Bisimulation through probabilistic testing. In Proceedings of the 16th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages.
[5] Rubner, Y.; Tomasi, C.; and Guibas L. J. 2000. The earth movers distance as a metric for image retrieval IJCV. 2000. 1