**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 \((S, A, R, F, \mathbb{P})\), a probabilistic bisimulation relation is an equivalence relation \(\sim\) on \(S\) such that if \(s_1 \sim s_2\) then

1. \(\forall s \in A, R_{s_1} = R_{s_2}\)
2. \(\forall s' \in S, F(s') = F(s')\)

\(\sim\) is defined as a partition \(\phi\) of \(S\) where

\[\phi(s) = \{s' \in S | s \sim s'\}\]

Let \(s, s'\) be two states, a bisimulation metric is a fixed point of \(F\) defined as follows

\[F(c(s)) = \max \{x | R_{s_1} = R_{s_2}, c(s') = c(s) + c(s')\}\]

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

**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 – 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 \(s, s'\) be two states, a bisimulation metric is an equivalence relation \(\sim\) on \(S\) such that if \(s_1 \sim s_2\) then

\[\forall s \in A, R_{s_1} = R_{s_2}\]

\(\sim\) is defined as a basis \(\phi\) on \(S\) where

\[\phi(s) = \{s' \in S | s \sim s'\}\]

**Complexity**

- The complexity of the update algorithm runs in \(O(|\phi|(|\phi||S| + |S|^2 \log |S|))\), which is an upper bound for the update at any step.

**Empirical Results**

**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.

**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. 1:** An example of probabilistic bisimulation relations.

**Fig. 2:** 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 4000, the number of features does not change substantially with the size of the space.

**Fig. 3:** Puddle World - computing the metric.

**Fig. 4:** Asynchronous computation: A plot of the approximation error in the value function computation \(\mathbb{P}(s, a)\) as the size of the alternative representation increases. This plot was generated on a Puddle World of size 1600.

**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.

**References**

2. Ferns, N., and Precup, D. 2014. Bisimulation metrics are optimal value functions. In UAI.