Bisimulation Metric Computation for Markov Decision Processes

[Extended Abstract]

Sherry Shanshan Ruan
McGill University
shanshan.ruan@mail.mcgill.com

Prakash Panangaden
McGill University
prakash@cs.mcgill.ca

Gheorghe Comanici
McGill University
gcoman@cs.mcgill.ca

Markov Decision Processes (MDPs) are a powerful mathematical model widely adopted in planning and learning under uncertainty. They have broad application in the area of artificial intelligence, computer vision, and operation research. A major task in the study of MDPs is to compute the optimal value function, which reflects a maximization of a cumulative measure of goodness of states in the long run. Dynamic programming algorithms such as value iteration are a typical means of computing the optimal value function [6]. However, in practice, the state space of MDPs is often too large to apply such dynamic programming algorithms. Recent research [1, 4] proposed to reduce the size of the state space using bisimulation-based techniques.

Bisimulation is an equivalence relation on the state space of a stochastic process [5]. Generally speaking, two states are bisimilar if one can simulate all the transitions of the other and the next state distributions are the same over bisimulation classes. Such a relation allows one to aggregate bisimilar states and thereby diminish the size of the state space while maintaining behavioral properties. But the notion of bisimulation relations is strict in the way that it aggregates only exactly bisimilar states. Ferns et al. [3] relax the notion of equivalence relations by introducing bisimulation metrics.

Bisimulation metrics can be regarded as a quantitative analogue of bisimulation relations. Metrics are superior to equivalence relations because they are smooth with respect to changes on the transition probabilities over the state space. Similar to bisimulation relations, bisimulation metrics can be used to cluster states. Alternatively, one can aggregate all states in ϵ-neighborhoods once a small parameter, ϵ, is carefully selected. In their work, Ferns et al. [3] provide an iteratively improving approximation algorithm and prove that the optimal value function of an aggregated MDP converges to the same value of the original MDP [3]. The metric allows one to quantify approximate aggregation strategies.

In spite of these desirable properties, computing bisimulation metrics is expensive. This is due to the fact that one has to compare all pairs of states at every iteration. Comanici et al. [2] tackle the problem by incorporating on-the-fly techniques which focus on a partial set of pairs of states. We extend their previous work by proposing an iterative asynchronous partition algorithm and provide important theoretical guarantees. Our work has four main contributions:

(1) Instead of computing the entire state space at every time step, we adopt heuristics to select certain pairs of states at each iteration and only update distances between them. The heuristic provides more flexibility in designing strategies for a variety of MDPs. We can develop suitable strategies for different MDPs based on real data.

(2) We use the Kantorovich metric to compare two probability distributions. The metric finds the optimal coupling of two probability distributions. We then adopt the Earth Mover's Distance (EMD) algorithms from computer vision to compute the similarity between states. By using our partition strategy, the algorithm only needs to search over a smaller set of couplings. Thus, the EMD algorithm becomes more tractable in finding the best coupling of probability distributions.

(3) We formalize the asynchronous partition algorithm in a mathematical framework and establish the correspondence between the theoretical characterization and the algorithm. Most importantly, we prove that our asynchronous partition algorithm converges to the same bisimulation metric provided that all the pair of states are selected infinitely often by the heuristic. Therefore, our formal framework bridges the gap between the previous work [2] and the theoretical warrants.

(4) We implement the described algorithm and apply it to synthetic data. The experimental results are consistent with the theoretical framework.

In future, we plan to extend experiments on synthetic data to real data with large state space using the proposed algorithm. Based on the experimental results, we expect to design more elaborate heuristics to further optimize the algorithm.

Keywords

Markov Decision Processes, bisimulation, Kantorovich metric

REFERENCES

[1] Marco Bernardo and Mario Bravetti. Performance measure sensitive congruences for markovian process algebras. *Theoretical Computer Science*, 290(1):117 – 160, 2003.

[2] G. Comanici, P. Panangaden, and D. Precup.

On-the-fly algorithms for bisimulation metrics. In *Quantitative Evaluation of Systems (QEST), 2012 Ninth International Conference on*, pages 94–103, Sept 2012.

[3] Norm Ferns, Prakash Panangaden, and Doina Precup. Metrics for finite markov decision processes. In Deborah L. McGuinness and George Ferguson, editors, *AAAI*, pages 950–951. AAAI Press / The MIT Press, 2004.

[4] Robert Givan, Thomas Dean, and Matthew Greig. Equivalence notions and model minimization in markov decision processes. *Artificial Intelligence*, 147(1 - 2):163 – 223, 2003. Planning with Uncertainty and Incomplete Information.

[5] Kim G. Larsen and Arne Skou. Bisimulation through probabilistic testing. *Information and Computation*, 94(1):1 – 28, 1991.

[6] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1994.