Reinforcement Learning with Dual Representations

Tao Wang, Michael Bowling, Dale Schuurmans

Abstract: We investigate the dual approach to dynamic programming and reinforcement learning, based on maintaining an explicit representation of stationary distributions as opposed to value functions. A significant advantage of the dual approach is that it allows one to exploit well developed techniques for representing, approximating and estimating probability distributions, without running the risks associated with divergent value function estimation. A second advantage is that some distinct algorithms for the average reward and discounted reward case in the primal become unified under the dual. In this paper, we present a modified dual of the standard linear program that guarantees a globally normalized state visit distribution is obtained. With this reformulation, we then derive novel dual forms of dynamic programming, including policy evaluation, policy iteration and value iteration. Moreover, we derive dual formulations of temporal difference learning to obtain new forms of Sarsa and Q learning. Finally, we scale these techniques up to large domains by introducing approximation, and develop new approximate off-policy learning algorithms that avoid the divergence problems associated with the primal approach. We show that the dual view yields a viable alternative to standard value function based techniques and opens new avenues for solving dynamic programming and reinforcement learning problems.