

Actor-Critic Algorithms and an Application to Vehicular Traffic Control

Shalabh Bhatnagar

Indian Institute of Science, Bangalore

Actor-critic algorithms are an important subclass of reinforcement learning methods. These are characterized by a parameterization of two entities – the actor and the critic. While the critic addresses the problem of prediction, the actor is concerned with control. Temporal difference (TD) learning is widely recognized as being an efficient method for the problem of prediction. By employing TD-based critics and actors that use policy or natural policy gradients, we present various actor-critic algorithms that boot strap in both the actor and the critic. Traditional reinforcement learning methods have been geared towards solving stochastic control problems that are usually modelled as Markov decision processes. We present an adaptation of one of the above algorithms with function approximation towards solving a Markov decision process with inequality constraints. We consider the long-run average cost criterion for both the cost and the constraint functions and employ the Lagrange multiplier approach for finding a constrained optimal policy. We sketch the convergence of our algorithms. Next, we present an application of one of our algorithms on the problem of traffic signal control. The problem here is to efficiently switch signals at traffic junctions by (a) adaptively finding the right order in which to switch and (b) finding the amount of time that a signal should be green, in order to maximize traffic flows and minimize delays. We observe that our algorithm significantly out-performs fixed timing algorithms as well as Q-learning which is a popular reinforcement learning algorithm.