POLICY GRADIENT METHODS IN MACHINE LEARNING

Jan Peters, University of Southern California, Los Angeles, USA, jrpeters@usc.edu
Evangelos Theodorou, University of Southern California, Los Angeles, USA, etheodor@usc.edu
Stefan Schaal, University of Southern California, Los Angeles, USA, sschaal@usc.edu

We present an in-depth survey of policy gradient methods as they are used in the machine learning community for optimizing parameterized, stochastic control policies in Markovian systems with respect to the expected reward. Despite having been developed separately in the reinforcement learning literature, policy gradient methods employ likelihood ratio gradient estimators as also suggested in the stochastic simulation optimization community. It is well-known that this approach to policy gradient estimation traditionally suffers from three drawbacks, i.e., large variance, a strong dependence on baseline functions and a inefficient gradient descent. In this talk, we will present a series of recent results which tackles each of these problems. The variance of the gradient estimation can be reduced significantly through recently introduced techniques such as optimal baselines, compatible function approximations and all-action gradients. However, as even the analytically obtainable policy gradients perform unnaturally slow, it required the step from 'vanilla' policy gradient methods towards natural policy gradients in order to overcome the inefficiency of the gradient descent. This development resulted into the Natural Actor-Critic architecture which can be shown to be very efficient in application to motor primitive learning for robotics.