8th International Conference on Applied Informatics
Eger, Hungary, January 27–30, 2010.

Using Gaussian Processes for variance reduction in Policy Gradient algorithms*

Jakab Hunor Sándor[a],

[a] Babes Bolyai University
e-mail:jakabh@cs.ubbcluj.ro

Abstract

Gradient based policy optimization algorithms (REINFORCE-like algorithms, Vanilla Policy Gradients, Natural Actor-Critic, Finite-difference methods) have many advantages over traditional value-function based methods when it comes to learning control polycies of complex robotic systems. However the majority of these methods suffer from high gradient variance which is a result of using Monte Carlo estimates of the Q-value function in the calculation of the gradient. $Q^{\pi}(x,a) \sim \left(\sum_{j=o}^{H} \gamma^j r_j\right)$ By replacing this estimate with a compatible function approximation on state-action space, the gradient variance can be reduced significantly. We propose the training of a Gaussian Process for the approximation of the action-value function $Q(\cdot,\cdot) \sim GP(m_q, k_q)$ which, after sufficient model confidence has been reached can be used to replace the Monte Carlo estimation. The learning of control polycies has to be performed online, because the only available data from which we can draw conclusions is provided by the agent's interaction with its environment. During the learning process at the end of each episode state-action pairs are selected and added to the set of support points for which the GP will be trained. At the gradient estimation stage we will choose between using the Monte Carlo estimates and the GP predictive mean, based on a measure which we define over the variance of the GP(the model confidence) at the current location in state-action space. We also investigate the possibility of using the information provided by the GP's generalization property to enhance our policy $\pi(x) = f(x,\theta) + \epsilon_{GP}$ by adding specific exploratory noise ϵ_{GP} to the parameterized controller output $f(x,\theta)$ which will bring the agent to more unexplored regions of the state-action space.

Jakab Hunor
Mihail Cogalniceanu str. 1. Cluj Napoca, Romania

*Thanks