Gaussian Process Policy Search with Latent Variables in Uncertain Environments

Name: Hikaru Sasaki
Laboratory's name: Robot Learning Lab.
Supervisor's name: Kenji Sugimoto

Abstract

Policy search reinforcement learning has been drawing much attention as a method for learning robot control. In particular, policy search using Gaussian process regression as the policy model can learn optimal actions from high-dimensional and redundant sensors as input. However, it is difficult to naively apply such a Gaussian Process Policy Search (GPPS) to real-world tasks because such tasks often involve various uncertainties. This is because the uncertainty of the environment often requires a very complex state-to-action mapping as a policy to obtain high-performance actions. To overcome such difficulties, this dissertation incorporates the concept of latent variable models in supervised/unsupervised learning. A latent variable model is a statistical model that relates a set of observable variables to a set of latent variables. If a latent variable model is appropriately designed, it may be possible to acquire a practical model that captures complex data while maintaining the simplicity of the learning algorithm. This dissertation explores a latent variable modeling approach in GPPS to cope with difficulty caused in uncertain environments. We then derive an algorithm that simultaneously performs latent variable inference and policy learning and aims to make policy search applicable to various real-world tasks with uncertainty. In particular, we addressed the following two issues: 1) multiple optimal actions emerging from a reward function with ambiguous specification, and 2) weak observations from the environment that contain little information about the state. We designed a policy model for each complexity by introducing latent variables into the Gaussian process and derived the policy update schemes based on variational Bayesian learning. The performance of the proposed policy search method was verified by simulation and a task using a robot manipulator. Finally, a policy learning framework based on Bayesian optimization with latent variables is proposed for application to actual heavy machinery, and its performance was verified using a real waste crane.