Learning in Robotics using Bayesian Nonparametrics

Marc Peter Deisenroth[1,2]
[1]Department of Computer Science
TU Darmstadt
Darmstadt, Germany

Dieter Fox[2]
[2]Department of CS&E
University of Washington
Seattle, WA 98195, USA

Carl Edward Rasmussen[3]
[3]Department of Engineering
University of Cambridge
Cambridge CB2 1PZ, UK

Abstract

We recently reported extraordinary successes for reinforcement learning in robotics and control using nonparametric Bayesian models [2, 3]. In particular, our approach needs only very little data to be collected from a physical system. Hence, it speeds up learning by an order of magnitude for a benchmark problem. This speed of learning allows us to apply our data efficient RL method to both robotic manipulators and relatively high-dimensional tasks that have not yet been learned from scratch. In this paper, we provide evidence that explicitly averaging out model uncertainties during planning and decision making is the key to success.

1 Bayesian Nonparametrics for Robot Learning

Reinforcement learning (RL) methods that learn from scratch *and* can be applied to robotic systems do hardly exist: It would require RL methods that succeed with very small data sets. The reason for RL's wide inapplicability to robotic systems is that model-free RL methods (they learn control strategies solely using samplings from the real system) require thousands or millions of "interactions" with the robot, which is often physically infeasible because robots can quickly wear out [3]. On the other hand, model-based learning methods (they learn a dynamics model based on a few samples and use this dynamics model to "emulate" the system to learn a control strategy) suffer from *model errors*, i.e., they assume that the learned model closely resembles the real physical system. In the case of only a few samples from the system, this assumption is often heavily violated [4, 1]. Therefore, the common approach is to collect more samples and then fit a regressor to more data points, which somewhat defeats the purpose of model-based learning.

Fig. 1 illustrates how model errors affect learning. Given a small data set of observed deterministic transitions (left), multiple transition functions plausibly could have generated the data (center). Choosing a single one causes severe consequences: When long-term predictions (or sampling trajectories from this model) leave the training data, the predictions of the function approximator are essentially arbitrary, but they are claimed with full confidence! By contrast, a probabilistic function

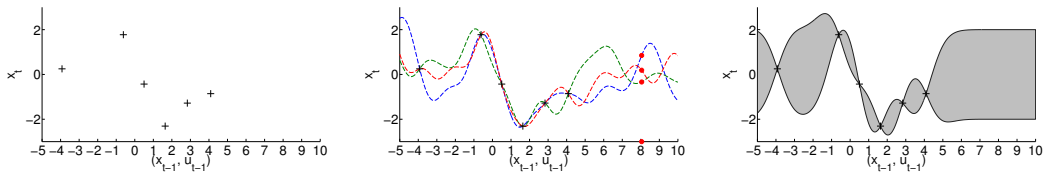

Figure 1: Small data set of observed transitions (left), multiple plausible deterministic (nonparametric) function approximators (center), probabilistic (nonparametric) function approximator (right). The probabilistic approximator models uncertainty about the latent function and allows for coherent predictions far away from the training data.

Table 1: Average learning success with nonparametric (NP) models.

	Bayesian NP model	Deterministic NP model
learning success	**94.52%**	0%

approximator places a posterior distribution over the transition function (right) and expresses the level of uncertainty about the model.

In [3], we proposed to use nonparametric Gaussian process (GP) models to learn distributions over dynamics models in the context of robotics. The GP posterior distribution allows us to express uncertainties about the underlying model in regions of the state space where we have no or only sparse data points. Using this GP dynamics model, we proposed an RL framework (policy search) called PILCO (probabilistic inference for learning control) that explicitly averages out the posterior GP model uncertainties during long-term planning and decision making, i.e., controller learning.

In [2], we showed on a common benchmark problem that our approach achieves an unprecedented speed of learning and learns at least an order of magnitude faster than other methods. In the following, we provide evidence that the Bayesian averaging over (posterior) model uncertainties is the key to successful learning from scratch.

2 Key to Success: Averaging over Posterior Model Uncertainty

In RL, we aim to find a controller parametrization θ^* that minimizes the expected long-term loss $J(\theta) = \sum_{t=0}^{T} \mathbb{E}[c(\mathbf{x}_t)]$, where c is an immediate cost function and $\mathbf{x}$ is the state of the system. The transition dynamics f are unknown. The state evolves according to a Markov process $\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t)$, where $\mathbf{u}_t = \pi(\mathbf{x}_t, \theta)$ is the control signal and π is the policy/controller. Given an initial distribution $p(\mathbf{x}_0)$ and a GP dynamics model (learned on previously collected data), PILCO finds a policy by computing the multiple-step ahead predictive distributions $p(\mathbf{x}_1), \ldots, p(\mathbf{x}_T)$ using approximate inference (moment matching), evaluating $\mathbb{E}[c(\mathbf{x}_t)]$ and, therefore, $J(\theta)$. Furthermore, PILCO computes the policy gradients $dJ/d\theta$ analytically [2], which are used within a CG optimizer to obtain θ^*.

In our experiments, we considered the standard benchmark cart-pole system. The system consists of a cart running on a track and a freely swinging pendulum. Using the control signal, the cart can be pushed to the left/right. Initially, the cart was expected to be in the center of the track with the pendulum hanging down. The goal was to learn a controller that swings the pendulum up and balances it in the inverted position. The continuous state space is 4D, and the continuous control space is 1D.

Assuming we employ a nonparametric model, we now evaluate whether Bayesian modeling is necessary to successful learning from scratch. To do so, we considered the PILCO learning framework [2, 3] using two different dynamics models: first, the standard GP model, second, a "deterministic GP" model, i.e., a GP where we consider only the posterior mean, but discard the posterior (model) variances during learning. Tab. 1 shows the average learning success of swinging the pendulum up and balancing it in the inverted position in the middle of the track. The average is taken over 8 learning setups with different random initializations. For each setup, PILCO had 15 iterations to learn the task. After that, we applied the learned controller 100 times to the system, where the start state was sampled from the initial state distribution $p(\mathbf{x}_0)$. Tab. 1 shows that learning is only successful when model uncertainties are taken into account during long-term planning and control learning.

3 Conclusion and Important Future Directions

Bayesian nonparametric models can have very practical and physically grounded applications in robotics/control. We provided evidence that *without* Bayesian modeling and inference, (reinforcement) learning from scratch does not succeed consistently or it requires intricate (parametric) prior knowledge about structure of the underlying model. To make Bayesian nonparametric models more widely accepted by the robotics/control community, we have to address the following issues: deal-

ing with large data sets (millions of data points, not thousands), fast approximate inference methods, increasing the interpretability of the learned models and how they relate to standard models, incorporation of domain knowledge. Perhaps above all, we need to demonstrate that Bayesian nonparametrics can be used to solve robotic/control *applications* better than traditional methods, where "better" can mean "more robust", "faster", or "with more general prior assumptions", for instance.

Acknowledgements

M. P. Deisenroth has been partially supported by ONR MURI grant N00014-09-1-1052 and by the EU project CompLACS.

References

[1] J. Andrew Bagnell and Jeff G. Schneider. Autonomous Helicopter Control using Reinforcement Learning Policy Search Methods. In *Proceedings of the International Conference on Robotics and Automation*, pages 1615–1620. IEEE Press, 2001.

[2] Marc P. Deisenroth and Carl E. Rasmussen. PILCO: A Model-Based and Data-Efficient Approach to Policy Search. In *Proceedings of the International Conference on Machine Learning*, pages 465–472, New York, NY, USA, June 2011. ACM.

[3] Marc P. Deisenroth, Carl E. Rasmussen, and Dieter Fox. Learning to Control a Low-Cost Manipulator using Data-Efficient Reinforcement Learning. In *Proceedings of the International Conference on Robotics: Science and Systems*, Los Angeles, CA, USA, June 2011.

[4] Jeff G. Schneider. Exploiting Model Uncertainty Estimates for Safe Dynamic Control Learning. In *Advances in Neural Information Processing Systems*. Morgan Kaufman Publishers, 1997.