

Active Learning for Robot Control

Philipp Robbel, MIT Media Lab, USA (robbel@mit.edu)

Sethu Vijayakumar, University of Edinburgh, UK (sethu.vijayakumar@ed.ac.uk)

Marc Toussaint, TU Berlin, Germany (mtoussai@cs.tu-berlin.de)

Abstract

The suggested poster summarizes and extends the approach presented in [1]. There, our work focused on learning the inverse dynamics of a robot arm using an efficient exploration strategy. We developed an active learning scheme for the LWPR algorithm [2] to guide data selection to regions of high model uncertainty until a task-specific trajectory can be realized by the manipulator. We position our work as an alternative to manual data collection (such as human guidance to specific points in the task space) and to simpler exploration schemes such as random flailing of the robot arm.

As in Vijayakumar et al's work, we use a compliant composite controller consisting of a learned feedforward model of the inverse dynamics and a low-gain corrective PID element. We address two difficulties with data selection for robot control: first, points cannot be chosen freely from the input distribution (order-sensitive scenario) and second, we would like to learn the inverse dynamics of the system online. Our active learning algorithm trades off between exploitation of the current LWPR model and exploration based on the confidence in the current LWPR predictions. In [1] we derive confidence bounds around the LWPR model and use these during learning as follows:

1. At every time step we determine model prediction and prediction confidence for the current query point \mathbf{x}_q . We postulate that the model generalization error is approximated by the size of the confidence intervals.
2. If the confidence is above a threshold, we apply the model prediction as a control signal to all joints and continue with step 1. Otherwise, we define \mathbf{x}_{q-1} , i.e., the last point that we trusted our model predictions, as a setpoint.
3. We execute a number of directed exploratory actions around \mathbf{x}_{q-1} to reduce the confidence interval size. In our current implementation those are random control signals followed by "resetting" the arm to the setpoint via high-gain PID control. We then continue with step 1.

By focusing exploration to the subspace where a particular task requires control, data collection of our algorithm is trajectory-specific. In [1] we presented results for a simulated 2-DOF robot arm along an arbitrary trajectory with a bell-shaped velocity profile. The work presented in this poster extends the evaluation to a simulated 6-DOF Puma arm. For both we demonstrate that our algorithm significantly reduces the number of data points (and time) required over random flailing to learn an inverse model that successfully drives back the trajectory. Evaluations are presented as reduction in normalized mean-square error (nMSE) along the trajectory with respect to the number of data points used during learning.

References

- [1] Philipp Robbel. Active learning in motor control. Master's thesis, University of Edinburgh, UK, August 2005. Available at <http://www.inf.ed.ac.uk/publications/thesis/online/IM050311.pdf>.
- [2] Sethu Vijayakumar, Aaron D'Souza, and Stefan Schaal. Incremental online learning in high dimensions. *Neural Computation*, 17(12):2602–2634, December 2005.