Bayesian Inference for Efficient Learning in Control

Marc Peter Deisenroth and Carl Edward Rasmussen

July 20, 2009

Motivation

In contrast to humans or animals, artificial learners often require more trials when learning motor control tasks solely based on experience. Efficient autonomous learners will reduce the amount of engineering required to solve control problems. By using probabilistic forward models, we can employ two key ingredients of biological learning systems to speed up artificial learning. We present a consistent and coherent Bayesian framework that allows for efficient autonomous experience-based learning. We demonstrate the success of our learning algorithm by applying it to challenging nonlinear control problems in simulation and in hardware.

Framework

For fast reinforcement learning, we require an adaptive probabilistic world model learned on previous experience, which we proposed in [**?**]. Due to limited experience, we utilize a probabilistic world model to represent uncertainty. The model is used for predicting and internal simulation to determine a good (model-optimized) policy based on the experience so far. However, due limited experience, it may be that the simulations do not correspond to the real world. When this situation occurs and we apply the model-optimized policy to the real system, the model will discover the discrepancy between the model's predictions and the world. This new insight will be incorporated into the subsequent model update.

Control Applications

Using Gaussian processes for the probabilistic world model and an arbitrary function approximator for the policy, our algorithm can naturally cope with continuous state and action spaces. We successfully applied it to control problems where only very general knowledge was available. In particular, no knowledge in form of a "teacher" has been used. We successfully applied our algorithm to the following tasks:

- Cart-pole: swing up and balancing (in computer simulation and hardware). Required experience: ≤ 1 minute. Figure 1 shows some snapshots of a test trajectory, where the controller is trained on experience from 17.5 s.
- Pendubot [**?**]: swing up and balancing (in computer simulation). Required experience: ≤ 3 minutes. Figure 2 sketches a solution to the Pendubot problem after an experience of about 75 s.
- Cart-double pendulum: (in computer simulation). Required experience: ≤ 3 minutes.

Videos of the learning tasks are provided at http://mlg.eng.cam.ac.uk/marc/learn_ctrl.

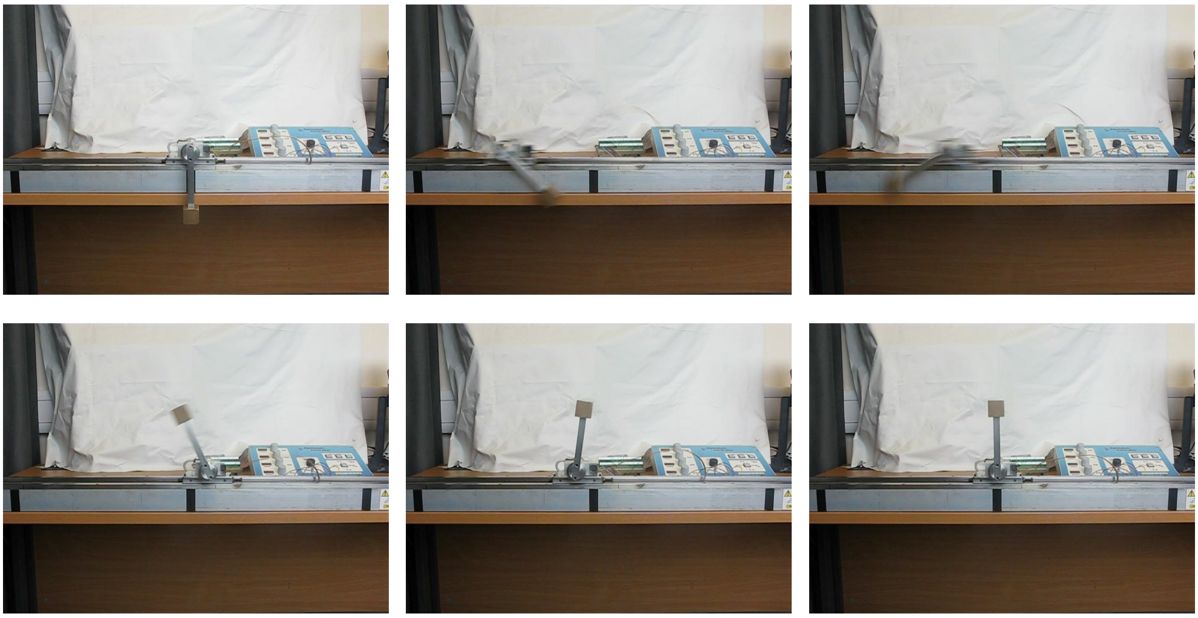

Figure 1: Performance of the underactuated cart-and-pole swing-up based on 17.5 seconds of interaction with the real system (from left to right and from top to bottom). Only the cart is actuated in order to swing the pendulum up. Once close to the upright position (green cross), the control law stabilizes the system around that state.

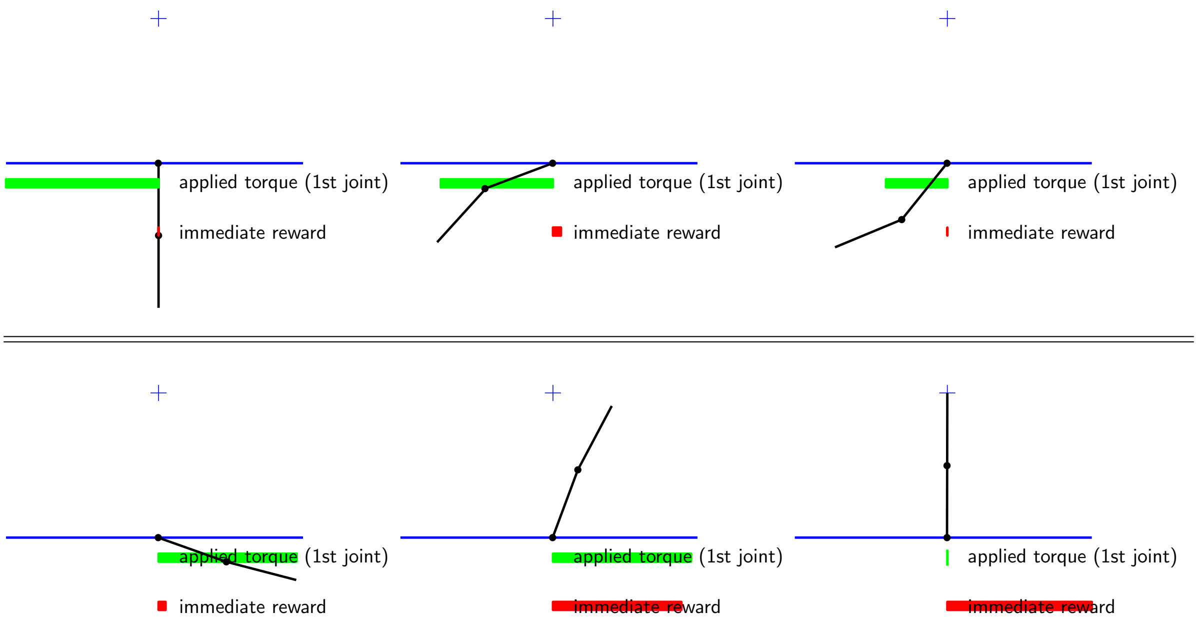

Figure 2: Pendubot swing-up and balancing based on less than two minutes' of interaction. Only the first joint is actuated in order to swing the double-pendulum up.