Sequencing Motor Primitives

Christian Daniel[1], Gerhard Neumann[1], Oliver Kroemer[1] and Jan Peters[1,2]

In recent years, policy search methods for robotics [1]–[4] have yielded encouraging results on tasks which are infeasible to encode by hand or to teach via demonstration. For example, Ng and Coates [3] have used policy search to develop autonomous stunt helicopters, Kober et al. [2] have shown how to learn the game 'ball in a cup', and Kormushev et al. [4] learned how to flip pancakes.

We focus on learning to sequence motor primitives with policy search. A motor primitive encodes an elemental movement and is typically represented as parametrized policy. Policy search methods directly search for parameters of the primitives that yield high rewards. While, the use of policy search methods is often limited to learning a single motor primitive, many complex tasks require the sequential combination of motor primitives. For example, playing a game of tennis does not only require a single hitting movement but a sequence of distinct hitting movements that finally result in scoring a point. Such a behavior requires strategic decisions on what type of motor primitive to use according to the current situation. Such primitives can also be sequenced in multiple ways to achieve a given task. Simultaneously representing a versatile strategic solution in the learned policy of the agent is desirable, as such knowledge typically improves the adaptability of the learned policy [5]. We present a policy search method that efficiently learns both the individual motor primitives and a versatile strategy to combine these primitives to achieve the long-term objective of the task.

Most policy search approaches are tailored for episodic policy search—the robot searches for a single parameter vector of the policy which is used throughout the whole episode. Hence, this setup allows the abstraction of a whole episode as a single decision: choosing the parameter vector of the policy at the start of the episode. Subsequently, the policy is executed with the specified parameters and the accumulated reward is observed by the agent. However, in order to multiple primitives sequentially, we need to use multiple parameter vectors sequentially throughout the episode. Learning such sequencing of motor primitives is mostly unsolved due to the high dimensionality of the problem.

In some approaches, the two problems of learning the motor primitives and sequencing the primitives are learned in isolation [6], [7]. However, such a strategy does not allow interaction between the two learning problems which might slow down learning or lead to an interference of both learning processes. Other approaches sequence primitives by using a combined policy parameter vector that contains the parameters of all primitives in sequence [8], [9]. Such strategy slows down learning as the parameter vector becomes unnecessarily high-dimensional. One approach for learning to sequence motor primitives in a single framework is given by Neumann et. al. [10]. However, this approach requires too many evaluations to be applied on a real robot.

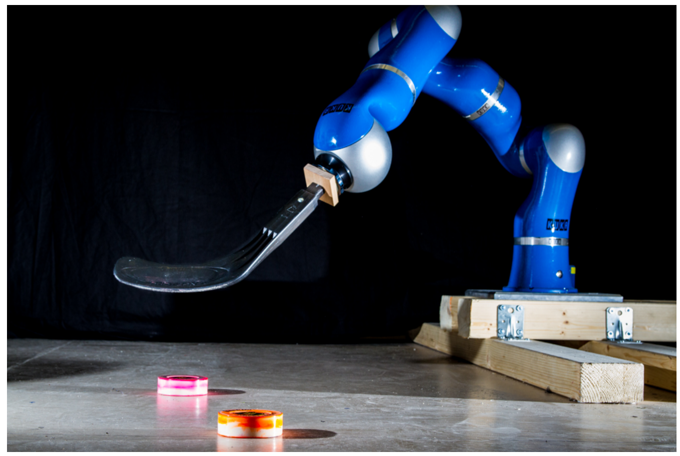

Fig. 1: The robot hockey task. The robot has to hit a puck into one of three target zones which yield different rewards. The puck can only be moved by shooting another puck at it. The robot has three shots to move the puck—overshooting the third zone yields zero reward.

We extend an existing episodic policy search method to learn the sequencing of multiple motor primitives while simultaneously improving the individual primitives. We base our algorithm on the Hierarchical Relative Entropy Policy Search (HiREPS) method [5]. HiREPS has two desirable properties for learning sequential motor primitives: it can adapt the movement primitives to the current situation and it allows the robot to learn versatile solutions for a single motor task. We extend the optimization problem defined by HiREPS to the finite-horizon case where each episode is composed of K motor primitives. The finite horizon formulation results in a time-indexed version of HiREPS, which allows us to learn individual policies for each decision time point. We are able to efficiently solve the temporal credit assignment problem by the use of additional constraints imposed by our finite horizon formulation. The policies at the single decision time points are connected by these constraints such that they jointly maximize the accumulated reward of the whole episode.

We evaluate our algorithm on robot hockey. This game requires learning a good strategy as well as individual motor primitives. Our integrated method can efficiently learn both elements of the problem simultaneously and outperforms the episodic counterpart of the algorithm.

[1]Technische Universität Darmstadt, Germany
[2]Max Planck Institute for Intelligent Systems, Tübingen, Germany
{daniel, neumann, kroemer, peters}@ias.tu-darmstadt.de

REFERENCES

[1] J. Peters and S. Schaal, "Natural Actor-Critic," *Neurocomputation*, vol. 71, no. 7-9, pp. 1180–1190, 2008.

[2] J. Kober, B. J. Mohler, and J. Peters, "Learning Perceptual Coupling for Motor Primitives," in *Intelligent Robots and Systems (IROS)*, 2008, pp. 834–839.

[3] A. Ng and A. Coates, "Autonomous Inverted Helicopter Flight via Reinforcement Learning," *Experimental Robotics IX*, 1998.

[4] P. Kormushev, S. Calinon, and D. G. Caldwell, "IEEE/RSJ International Conference on Intelligent Robots and Systems," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2010.

[5] C. Daniel, G. Neumann, and J. Peters, "Hierarchical Relative Entropy Policy Search," in *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2012.

[6] J. Morimoto and K. Doya, "Reinforcement Learning of a Dynamic Motor Sequence: Learning to Stand-Up," in *International Conference on Intelligent Robots and Systems (IROS)*, 1998.

[7] J. Kober and J. Peters, "Learning Elementary Movements jointly with a Higher Level Task," in *Intelligent Robots and Systems (IROS)*, 2011, pp. 338–343.

[8] C. Daniel, G. Neumann, and J. Peters, "Learning Concurrent Motor Skills in Versatile Solution Spaces," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012.

[9] F. Stulp and S. Schaal, "Hierarchical Reinforcement Learning with Movement Primitives," in *2012 IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2011, pp. 231–238.

[10] G. Neumann, W. Maass, and J. Peters, "Learning Complex Motions by Sequencing Simpler Motion Templates," in *International Conference on Machine Learning (ICML)*, ser. (ICML 2009), 2009.