Active body schema learning

Ruben Martinez-Cantin, Manuel Lopes and Luis Montesano

Abstract

Humanoid and industrial robots are becoming increasingly complex. Most of the algorithms to control these systems require, or are improved with, a full kinematic model of the system. In this work, we are interested in autonomously learning the kinematic description, a.k.a. body schema, for unknown systems. Even for a calibrated system, the ability to continuously tune its parameters allows the system to cope with failures, wear and tear and modifications of the kinematic chain. A natural inspiration for this type of learning is biology. Animals have knowledge about its own body allowing them to perform a large variety of motor tasks. Even after an injury, including the extreme case of limb removal, they can adapt their actions and continue to act [1].

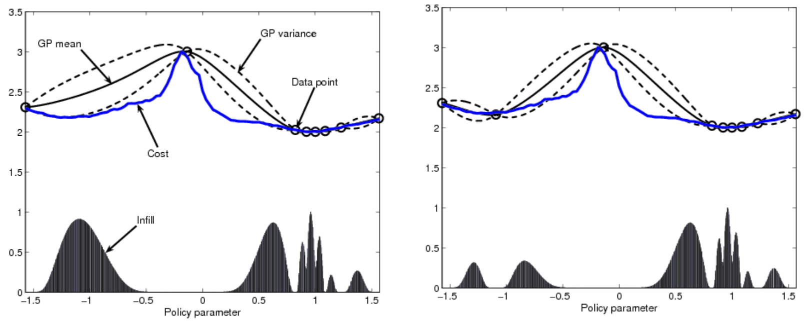
Inspired by the way humans develop, previous work on body schema learning has focused on computing the kinematic model based on random motions [2, 3]. Most of these works require quite a lot of data to converge to the right solution and pointed out that active strategies can help to reduce this complexity. In this paper, we study an active approach to the body schema learning. The main idea is to select motions that are more informative with respect to the current knowledge about the body schema.

To apply an active strategy for learning, the robot is described using a physical parametric model of the body schema (e.g. the joint locations and orientations of a robotic arm). Learning the body schema is then a parameter estimation problem. We adopt a Bayesian perspective and compute the posterior of these parameters based on observations of the end-effector of the arm and the motion commands (arm configurations). The distribution over the parameters is sequentially updated using Recursive Least Squares (RLS) when new observations are gathered.

Having this posterior, it is possible to formulate an optimal exploration strategy. Unfortunately this results in an NP-complete problem. To overcome this limitation we use model predictive control, i.e. a fixed finite horizon, and perform a policy search over this horizon. This search on the space of policies is again a complex procedure, and we use an anytime adaptive online algorithm [4] to select the best policy at every time step. Standard tehcniques such as linear-quadratic-Gaussian (LQG) controllers are not directly applicable because the model is nonlinear and non-Gaussian and the cost function is not quadratic. Also, since the action and parameter spaces are large-dimensional and continuous, one cannot use methods based on discretization of the problem [5].

Our approach is based on an global optimization method using a response surface of the expected cost of the actions. This surface is modeled as a Gaussian Process (GP), which allows to design exploration strategies in an *intelligent* way, i.e., based on Bayesian design of new queries. Also, in this work, we introduce some improvements on the algorithm from [4] that reduce the standard GP prediction cost $\mathcal{O}(n^3)$ to $\mathcal{O}(n^2)$.

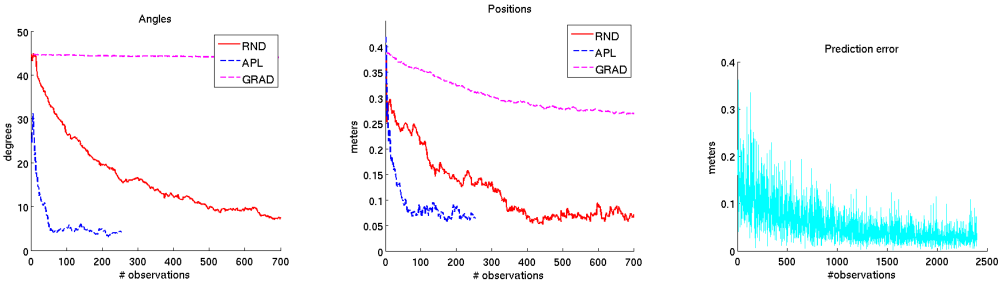
Intuitively, the process is as follows. To find a new exploration point, we start randomly sampling a small set of policies (robot configurations) and computing their expected cost. Based on this values, we



(a)

(b)

Figure 1: An example of active policy learning with a univariate policy using data generated by our simulator. (a) The cost function is approximated using 11 data points. The figure also shows the GP estimate and the infill function of each potential query. The infill is high where the GP predicts a low expected cost (exploitation) and where the prediction uncertainty is high (exploration). (b) Same function after selecting and labeling the query with the highest infill. The new infill function suggests that, for the next iteration, we should query a point where the cost is expected to be low (exploitation).



(a) (b) (c)

Figure 2: Convergence of the body schema learning for Baltazar with 6 dof. (a) mean error in the relative orientations of the joint angles, (b) mean error in the location of the joint angle and (c) prediction error during the learning. Each figure shows the results for random exploration using the RLS (rnd), the active approach (apl) and an stochastic gradient method (grad). (c) Prediction error for the gradient method.

initialize the Gaussian process to approximate the cost function. Then, we use an active learning strategy to explore the policy space (robot configuration space) to improve the cost function approximation to find its minimum. By using an Infill function [6], which trades-off the exploration and exploitation in a Bayesian way, the exploration is concentrated in those areas of the space that have higher probability of having a lower cost. Basically the infill function computes a value which is a weighted combination of the regression prediction and its uncertainty. Figure 1 provides a simple 1D example of this intuition and illustrates the infill function.

Note that the resulting algorithm applies active learning at two levels. The first one aims to find the best configuration. To make such a decision, one needs to explore the policy space. This is also done in an active way through internal simulations instead of using random or systematic sampling of the policies.

The proposed algorithm has been evaluated on simulation, up to 12-dof, and on a real 6-dof humanoid robot. We compared our approach to gradient methods, which are the state of the art for that problem. The results show that the proposed policy search greatly reduces the number of observations required for the same error level in an order of magnitude. As a by product, the proposed Bayesian estimator seems to be more robust to initializations and to achieve better solutions in average as can be seen in figure 2.

Acknowledgments:
Work partially supported by the ICTI and FCT, under the CMU-Portugal Program, the (POS_C) program that includes FEDER funds and the projects PTDC/EEA-ACR/70174/2006, and (IST-004370) RobotCub.

References

[1] P. Haggard and D. Wolpert, "Disorders of body scheme," in *Higher-Order Motor Disorders*, H. Freund, M. Jeannerod, M. Hallett, and R. Leiguarda, Eds. Oxford University Press, 2005.

[2] J. Sturm, C. Plagemann, and W. Burgard, "Adaptive body scheme models for robust robotic manipulation," in *RSS - Robotics Science and Systems IV*, Zurich, Switzerland, june 2008.

[3] M. Hersch, E. Sauser, and A. Billard, "Online learning of the body schema," *International Journal of Humanoid Robotics*, vol. 5, no. 2, pp. 161–181, 2008.

[4] R. Martinez-Cantin, N. de Freitas, and J. Castellanos, "Active policy learning for robot planning and exploration under uncertainty," in *Proc. of Robotics: Science and Systems*, 2007.

[5] R. Smallwood and E. Sondik, "The optimal control of partially observable Markov processes over a finite horizon," *Operations Research*, vol. 21, pp. 1071–1088, 1973.

[6] J. Mockus, V. Tiesis, and A. Zilinskas, "The application of Bayesian methods for seeking the extremum," in *Towards Global Optimisation 2*, L. Dixon and G. Szego, Eds. Elsevier, 1978, pp. 117–129.