A machine learning approach to reaching tasks

D. MARIN*† and O. SIGAUD†

† Institut des Systèmes Intelligents et de Robotique – CNRS UMR 7222

UPMC – 4 place Jussieu 75252 Paris Cedex 05 France

Keywords: human motor control; reaching; optimal control; optimization

1 Introduction

Optimal Control (OC) is a useful framework for modeling Human Motor Control (HMC) properties [5], [1], but the corresponding methods are too expensive to be applied on-line and in real-time as would be required for modeling everyday movements. A straightforward solution to this cost problem consists in improving a parametric feedback controller all along the lifetime of the system through its interactions with its environment. Such a method also comes with the benefits of adaptation and can be implemented by using incremental, stochastic optimization methods.

In [3], a model of human reaching movement was proposed based on the assumption that HMC is governed by an optimal feedback policy computed at each visited state given a cost function. This model explains the optimal movement time as emerging from a trade-off between the utility of successfully reaching and the cost in terms of efforts. In this paper, we present a machine learning approach to get a reactive parametric controller that performs well w.r.t. their method and can be further improved through stochastic optimization, while being fast enough to be used on-line. Moreover, it shows interesting generalization capabilities.

2 Methods

Our approach consists in 3 steps: first, the Near-Optimal Planning System (NOPS) of [3] is used to generate a few sample trajectories, i.e. sequences of state-action pairs. Then, a state-of-art regression method, XCSF [6], is fed with these pairs to learn a mapping from states to actions. As XCSF relies on a sum of weighted local models, it can be to used as a parametric controller, called "XCSF controller". This controller will have good initial parameters thanks to NOPS demonstrations, while being able to compute its control in real-time. Finally, a stochastic optimization can be performed over these parameters using a direct Policy Search method based on a Cross Entropy method, CEPS [2], which robustly improves the controller performance by trial and errors.

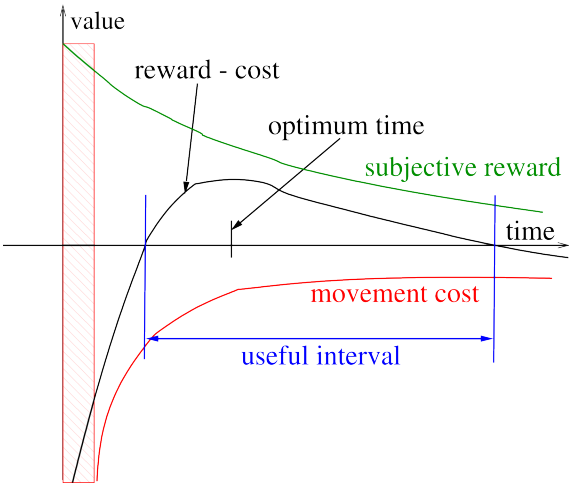

Fig. 1. Optimal movement time. Reaching cannot be performed under a certain time (dashed area) and is less and less costly in terms of efforts as the movement is performed more slowly (red line). However the subjective reward for reaching the goal decreases with time to account for greediness, i.e., we are less interested in gains that will occur in a distant future than at the present time. The reward versus cost criterion, resulting from the sum of the subjective reward and the (negative) cost reaches a maximum for a certain time. When the criterion is negative (outside useful interval), the subject should not move, i.e., the movement is not worth it. Details can be found in [3], [2].

3 Results and Discussion

We apply our approach to a simulated reaching task with a 2 degrees-of-freedom planar arm actuated by 6 muscles (Fig. 2, see [2] for details).

We demonstrate the generalization capabilities of our approach using NOPS demonstrations for a small set of targets, located in the central region of

***Corresponding author. Email: marin@isir.upmc.fr**

the workspace, and testing the XCSF controller on a larger set that covers the workspace. The performance that NOPS optimize, i.e. the reward versus cost criterion illustrated in Fig. 1, is used for evaluating all controllers. Table 1 shows that the XCSF controller performance over the small set is very close to NOPS, but decreases as targets become more distant, resulting in a poor mean performance over the large set. However, it can be significantly improved by allowing CEPS to perform trial and errors for targets of the large set. The performance over the small set slightly decreases as a consequence of generalization.

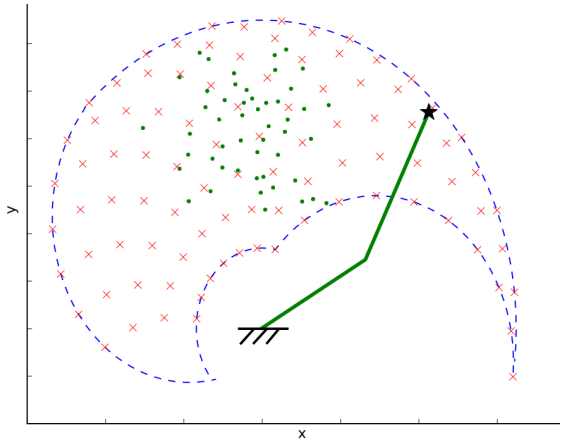

Fig. 2. The arm workspace. The reachable space is delimited by a dashed line envelope. Arm segments are represented by bold green lines. Start position is represented as a star, small target set as green dots and large target set as red crosses.

	small set	large set
NOPS	28.22 ± 2.06	27.46 ± 3.73
XCSF	**27.45 ± 2.35**	2.44 ± 46.03
XCSF+CEPS	22.96 ± 11.50	**12.76 ± 22.28**

Table 1. Mean perfomance over each target set and for each method. Sets used for learning are in bold.

4 Conclusions

We have presented a machine learning approach to modelling HMC applied to reaching task. As in [3], the speed of movement emerges from a compromise between the subjective value of the reward and the cost of movement. As opposed to The computational cost of OC methods is avoided by optimizing a reactive parametric controller, which offers generalization capabilities that makes the learning process reasonably easy in practice.

From a neurosciences perspective, this parametric approach might be seen as a computational model of how the Central Nervous System might store the capability to reach optimally and in real-time from any start position and to any target.

For now, our approach requires knowledge of the exact arm dynamics, as it is needed by NOPS, however we could replace it by a learned model [4]. On the long run, we would like to apply such approach to the whole body control of humanoid robots like iCub, so that they perform more human-like movements.

References

[1] E. Guigon, P. Baraduc, and M. Desmurget, Optimality, stochasticity and variability in motor behavior. *Journal of Computational Neuroscience*, 24(1):57–68, 2008.

[2] D. Marin, J. Decock, L. Rigoux, and O. Sigaud, Learning cost-efficient control policies with XCSF: generalization capabilities and further improvement. In *Proceedings of the 13th annual Conference on Genetic and Evolutionary Computation (GECCO)*, 1235–1242, 2011.

[3] L. Rigoux, O. Sigaud, A. Terekhov, and E. Guigon, Movement duration as an emergent property of reward directed motor control. In *Proceedings of the Annual Symposium Advances in Computational Motor Control*, 2010.

[4] O. Sigaud, C. Salaun and V. Padois. On-line regression algorithms for learning mechanical models of robots: a survey. *Robotics and Autonomous Systems*, 59(12), 1115–1129, 2011

[5] E. Todorov, Optimality principles in sensorimotor control. *Nature Neurosciences*, 7(9):907–915, 2004.

[6] S. W. Wilson, Classifiers that Approximate Functions. *Natural Computing*, 1(2-3):211–234, 2002.

Acknowledgments

This work was supported by the Ambient Assisted Living Joint Programme of the European Union and the National Innovation Office (DOMEO-AAL-2008-1-159), more at http://www.aal-domeo.eu.