Bottom-Up Skill Building for Bold Hearts 3D 2007

D. Polani, J. Garner, S. Jayasudha, P. Haghighi Rad,
M. Snelling, P. Moe, and F. Yeoung

School of Computer Science
University of Hertfordshire, UK

Abstract. Consistent with our earlier work and the other work in our research group, we are interested in developing principled methods to learn behaviours. This is particularly interesting and challenging in humanoids since any possible tactical, strategical or cooperative aspects can only be successfully tackled once the basic skills are in place.
The construction of basic skills in humanoids is usually an intricate business that requires a large amount of hand-tuning. We aim to develop a path towards reducing this amount of handtuning and moving towards a "self-organized" learning methodology.
The approach championed here in skill learning is based on biologically inspired algorithms (such as GAs, and central pattern generators), guided by information-theoretic quantities which help to structure the search space and to search it more efficiently.

1 Introduction

Consistent with the work of the last years, and with the other research goals in our group, the main goal of this year's team is to move towards a comprehensive bottom-up strategy for skill development and learning. Doing so is particularly challenging, interesting and relevant in humanoids since any possible tactical, strategical or cooperative aspects can only be successfully addressed, once the basic skills are in place. We particular wish to address the fact that the construction of basic skills in humanoids is usually an intricate business that requires a large amount of hand-tuning. While interesting from a purely engineering point of view, from an Artificial Intelligence point of view, it is unsatisfactory that such skills are essentially to be designed by hand, rather than being developed in a bottom-up approach. An additional incentive to study the bottom-up view is provided by biology: we know that organisms are very good at learning skills (or even evolving them). In particular, if the environment or the agent itself changes, it would be desirable if it were not difficult to adapt the agent to the new situation, as in biology. Thus, our goal is to develop a path towards reducing this amount of handtuning and moving towards a "self-organized" learning methodology. We have been moving towards this path in the last years, but the relative ease with which good control for the *sphere* robots can be designed, makes the effort of a systematic design of behaviour sets for the simple sphere robots obsolete — hand-crafted solutions fare quite strongly, with explicitly designed rules. However, we have the strong expectation that the requirements for the development of strong and general humanoid control skills has the right level of complexity to be sufficiently challenging to make the automatic approach relevant and ultimately competitive.

The approach we champion here in skill learning is based on biologically inspired algorithms (such as GAs, and central pattern generators), but who are in the same time guided by information-theoretic quantities which help to structure the search space and to search it more efficiently.

2 Evolution of Walk Patterns (Demo Executable)

We have begun with experiments for the evolution of walk patterns. Due to the limited time the server was available, only short evolutionary runs of a few hours length could have been performed until now. The submitted executable shows some of the evolved walk patterns for a multiobjective optimization[1] according to two objectives: walk time (an agent is killed if its head — approximately — drops below a certain point), and walk distance (the length of the path between the start of the agent to the moment when it is killed). The idea is to evolve the parameters of the walk pattern (amplitude, frequency and phase) for the legs. The feet (parameters 5 and 6) are frozen, and such is the lateral leg movement (parameter 3), similarly the arms are frozen in these experiments. The file `oscparam.dat` needs to be copied into the current directory where the executable (the script `bh_start` makes sure that the `LD_LIBRARY_PATH` covers the special dynamic libraries) is placed. It should be started several times to appreciate the results in their variation (in the current version, the agent is blind — it only uses its sensors to report back the length of the walk to the Genetic Algorithm). Also, some different `oscparam<xxx>.dat`-files are given, and should be copied onto `oscparam.dat`, to try out different walk patterns. Currently, the stability is still a problem, but, as remarked, we expect much longer runs (at least 24-48 hours) to give more satisfactory results. One advantage of using a Genetic Algorithm is the fact that one can hope to discover stable walk patterns which are not exactly human-plausible (e.g. the "shaky" patterns of the agent — see for instance `oscparam_7*.dat` provide a comparatively long-lived walking pattern).

3 Guidance through Information

We emphasized in earlier team descriptions that Shannon information can be a powerful indicator of where "interesting" properties of the world lie. This is important because, although learning, specifically *reinforcement learning* has been part of the RoboCup endeavour for a longer time [21, 19, 2, 16]. As emphasized in

[1] As multiobjective GA, we use NSGA II as described in [4], with the code provided by the authors.

[7], reinforcement learning methods are of interest because of their generality and mathematical grounding. They are also quite successful in nontrivial problems [20]; in conjunction with kernel methods, they can address even larger problems in a highly efficient way [6, 12, 11, 10].

Still, the problems to address are quite large (and large-dimensional), however, realistic embodied agents offer a selection of possible partial decompositions [9]. The use of information-theoretic (or information-theoretically motivated) decompositions is a natural approach. The complexity of the sensor/actuator space for the agents at question suggests either the use of an a-priori pre-structuring la informatory sensoritopic map [17], or some kind of projection of the sensory state map onto a lower-dimensional (and informationally more parsimonious) manifold according to the Optimal Manifold Algorithm [3]. Another direction, which we will emphasize this year, is the use of *empowerment* [14, 15], an information-theoretic quantity which identifies “interesting” areas in the sensorimotor space, and is defined through the *potential information flow*[2] through the environment using the agent’s sensorimotor loop. We have studied this quantity in a number of sensorimotor scenarios, and found that it is highly plausible as a possible candidate for taskless utilities (other examples include the novelty detection [13], homeokinesis [5], excess entropy [18]). Here, however, we will stick to the empowerment measure as developed by our group, because of its well-understood ramifications and connections with the other issues concerning information-theoretical views onto the perception-action loop of embodied agents. We expect empowerment to provide us with *local* learning utilities which will allow us to learn motoric skills incrementally — it will be used to direct the GA (and possibly other, more sophisticated learning algorithms) towards *interesting* areas of behaviour, before requiring to solve a large-scale problem as a whole[3].

References

1. Ay, N. and Polani, D. Information Flows in Causal Networks. Proc. NIPS Workshop on Causality and Feature Selection, Dec 2006.
2. Buck, S. and Riedmiller, M. Learning situation dependent sucess rates of actions in a robocup scenario. In *Proceedings of PRICAI ’00, Melbourne, Australia, 28.8.-3.9.2000*, page 809, 2000.
3. Chigirev, D. and Bialek, W. Optimal manifold representation of data: An information theoretic perspective. In *Advances in Neural Information Processing (NIPS 16)*, Cambridge, 2004. MIT Press.
4. Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6:182–197, 2002.
5. Der, R. Selforganized robot behavior from the principle of homeokinesis. In Groß, H.-M., Debes, K., and Böhme, H.-J., editors, *Proc. Workhop SOAVE ’2000 (Selbstorganisation von adaptivem Verhalten)*, volume 643 of *Fortschritt-Berichte VDI, Reihe 10*, pages 39–46, Ilmenau, 2000. VDI Verlag.

[2] For a definition of information flow, see [1].

[3] Quite recently, the plausibility of successfully achieving hierarchical learning has received renewed attention by a hierarchical learning algorithm [8].

6. Engel, Y., Mannor, S., and Meir, R. Bayes meets Bellman: The Gaussian Process Approach to Temporal Difference Learning. In *Proc. of ICML 20*, pages 154–161, 2003.
7. Franco, S. and Polani, D. Skill Learning Via Information-Theoretical Decomposition of Behaviour Features. In Polani, D., Browning, B., Bonarini, A., and Yoshida, K., editors, *RoboCup 2003: Robot Soccer World Cup VII*, volume 3020 of *LNCS*. Springer, 2004. Team Description (CD supplement).
8. Hinton, G. E. and Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 28. July 2006 2006.
9. Jacob, D., Polani, D., and Nehaniv, C. L. Legs that can walk: Embodiment-Based Modular Reinforcement Learning applied. In *IEEE Computational Intelligence in Robotics & Automata (IEEE CIRA 2005)*, pages 365–372. IEEE, 2005.
10. Jung, T. and Polani, D. Sequential Learning with LS-SVM for Large-Scale Data Sets. In *Proc. 16th International Conference on Artificial Neural Networks, 10-14. September 2006, Athens, Greece*, volume 2, pages 381–390, 2006.
11. Jung, T. and Polani, D. Kernelizing LSPE(λ). In *Proc. 2007 IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning, April 1-5, Hawaii*, 2007. Accepted 2. December 2006. In Press.
12. Jung, T. and Polani, D. Learning RoboCup-Keepaway with Kernels. In Lawrence, N., Schwaighofer, A., and Candela, J. Q., editors, *Gaussian Processes in Practice*, volume 1 of *JMLR Workshop and Conference Proceedings*, pages 33–57, 2007. Accepted 30. January 2007.
13. Kaplan, F. and Oudeyer, P.-Y. Maximizing learning progress: an internal reward system for development. In Iida, F., Pfeifer, R., Steels, L., and Kuniyoshi, Y., editors, *Embodied Artificial Intelligence*, volume 3139 of *LNAI*, pages 259–270. Springer, 2004.
14. Klyubin, A. S., Polani, D., and Nehaniv, C. L. All Else Being Equal Be Empowered. In *Advances in Artificial Life, European Conference on Artificial Life (ECAL 2005)*, volume 3630 of *LNAI*, pages 744–753. Springer, 2005.
15. Klyubin, A. S., Polani, D., and Nehaniv, C. L. Empowerment: A Universal Agent-Centric Measure of Control. In *Proc. IEEE Congress on Evolutionary Computation, 2-5 September 2005, Edinburgh, Scotland (CEC 2005)*, pages 128–135. IEEE, 2005.
16. Lauer, M. and Riedmiller, M. An Algorithm for Distributed Reinforcement Learning in Cooperative Multi-Agent Systems. In *Proc. 17th International Conf. on Machine Learning*, pages 535–542. Morgan Kaufmann, San Francisco, CA, 2000.
17. Olsson, L., Nehaniv, C. L., and Polani, D. From Unknown Sensors and Actuators to Actions Grounded in Sensorimotor Perceptions. *Connection Science*, 18(2):121–144, 2006. Special Issue on Developmental Robotics, Douglas Blank and Lisa Meeden, editors. [ISSN: 0954-0091, Online ISSN: 1360-0494].
18. Prokopenko, M., Gerasimov, V., and Tanev, I. Measuring Spatiotemporal Coordination in a Modular Robotic System. In Rocha, L. M., Bedau, M., Floreano, D., Goldstone, R., Vespignani, A., and Yaeger, L., editors, *Proc. Artificial Life X*, August 2006. [ISBN-10: 0-262-68162-5, ISBN-13: 978-0-262-68162-9].
19. Stone, P. *Layered Learning in Multiagent Systems: A Winning Approach to Robotic Soccer.* MIT Press, 2000.
20. Stone, P., Sutton, R. S., and Kuhlmann, G. Reinforcement Learning for RoboCup-Soccer Keepaway. *Adaptive Behavior*, 13(3):165–188, 2005.
21. Stone, P. and Veloso, M. A layered approach to learning client behaviors in the RoboCup soccer server. *Applied Artificial Intelligence*, 12, 1998.