A neural pre- and post-processing framework for goal directed behavior in self-organizing robots

Frank Hesse (P)[1,2], Poramate Manoonpong[1,2] and Florentin Wörgötter[1,2]

[1] Third Physics Institute - Biophysics, Georg-August-Universität Göttingen
[2] Bernstein Center for Computational Neuroscience, Göttingen
E-mail: f.hesse@physik3.gwdg.de

Abstract—In the work at hand we introduce a neural pre- and post-processing framework whose parameters can be adapted by any learning mechanism, e.g. reinforcement learning. The framework allows to generate goal-directed behaviors while at the same time exploiting the beneficial properties, e.g. robustness, of self-organization based primitive behaviors.

Keywords—Neural Networks, Self-Organization, Reinforcement Learning, Autonomous Robots

1 Introduction

The definition of preprogrammed (primitive) behaviors for autonomous robots requires an understanding of the robotic system in order to actuate it properly [1]. To minimize these efforts self-organizing control generating such behaviors seems promising. In this work, we investigate how self-organization of motor primitives and learning of goal-oriented behaviors can be combined. Therefore, we introduce a general neural pre- and post-processing framework (sec. 3), in which basic motor primitives generated by a first controller (H1) can be modified by a second controller (H2) in order to achieve a desired behavior or a given goal. Since only the sensor values and motor commands but no internal parameters of H1 are modulated the principle can be applied to basically every controller.

In the work at hand, H1 is a self-organizing controller (sec. 2), hence cannot generate goal oriented behaviors. H2 is based on reinforcement learning and can shape the generated behaviors in order to achieve a given goal (sec. 4).

2 Self-organizing Control

For the self-organizing control of the robotic device the homeokinetic principle [2] is employed. This principle does not take externally specified goals or reference values into account. Instead it generates various primitive behaviors in autonomous robots, depending on the properties of robot and environment. From the homoekinetic principle, learning rules for neurons of closed loop robot controllers can be derived.

3 Neuronal Pre- and Post-Processing Framework

Let us consider a robot that provides at each instant of time, $t = 0, 1, \ldots$, a vector of sensor values $x_t^R \in \mathbb{R}^n$ and is controlled by the motor values $y_t^R \in \mathbb{R}^m$. With $n \geq m$ the robot is assumed to not only provide (proprioceptive) feedback about the actuators but maybe also context information required for the task at hand. The controller H1 generates in each time step a vector of motor commands $y_t^C \in \mathbb{R}^k$ (with $k = m$) based on the sensory input $x_t^C \in \mathbb{R}^l$ (with $n \geq l \geq m$), where usually only the feedback about the actuators is used.

The post processing is an adaptable mapping $M : \mathbb{R}^k \to \mathbb{R}^k$, which depends on the parameters $b \in \mathbb{R}^k$ defining the amplitude change of the motor command of H1 before it is executed by the robotic hardware: $y^R = M(y^C)$ with

$$y^{R,i} = b^i y^{C,i}, \quad b^i > 0 \text{ for all actuators } i. \tag{1}$$

The preprocessing is the inverse mapping $M^{-1} : \mathbb{R}^l \to \mathbb{R}^l$ where the sensor values of the robot are scaled inversely to the motor command[1]:

$$x^{C,i} = \frac{1}{b^i} * x^{R,i}, \quad i = 1, ..., l \tag{2}$$

The adaptation of the parameters b is done by the controller H2 (sec. 4). The pre and post processing can be represented as artificial neurons with linear output function whose input weights are $\frac{1}{b}$ and b respectively, see figure 1.

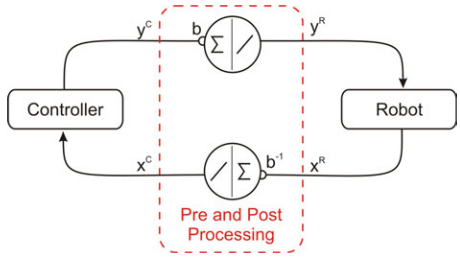

Figure 1: Neuronal representation of the proposed pre- and post-processing framework for one degree of freedom.

Assuming an ideal actuator which can immediately execute motor commands ($x_t^{R,i} = y_t^{R,i}$) and together with equations 1, 2 we find $x^{C,i} = \frac{1}{b^i} * b^i * y^{C,i} = y^{C,i}$. Hence from the point of view of the self-organizing controller (H1) there is no influence caused by the pre and post processing.

4 Goal-oriented control

The goal-oriented control (H2) is used to adapt the parameters of the neural pre and post processing in order to achieve a given goal. In principle any method could be used. In order to emphasize the presented framework and not the details of a complex learning mechanism a simple reinforcement learning method, Q-Learning [3] with a greedy behavior policy, is used.

[1] In the case of $l > k$ (more sensors then motors) senors connected to the same motor share the same b.

5 Experiments

The experiments with a four-wheeled robot in a square arena with four goals (cp. figure 2) were conducted in the physics simulator LpzRobots [4].

The homeokinetic controller (H1), which has to coordinate the wheels in order to bring the robot into motion, is provided with the preprocessed measured wheel velocities $x_t^{C,i} = \frac{1}{b^i} x_t^{R,i}, \quad i = 1, \ldots, 4$, as sensory input. The motor command $y_t^C \in \mathbb{R}^4$ of H1 is passed to the robot via the post processing: $y_t^{R,i} = b^i y_t^{C,i}, \quad i = 1, \ldots, 4$.

The goal-oriented controller H2 guides the robot to the goal positions in the environment, which are selected in serial order. Sensor information is the angle between direction of motion and currently active goal location. The actions of the Q-learning correspond to changes of the parameters $b^i, \quad i = 1, \ldots, 4$, allowing to steer the robot. In the first minute only H1 was active, while during the following ten minutes H2 was learned. After learning two tasks had to be executed.

Also pure homeokinetic control (no pre or post processing) and the "common" control paradigm with manually predefined instead of self-organization based motor primitives were investigated. The first task (*normal* case) for the learned systems is to let the robot follow a sequence of goal positions for ten minutes. A new goal is activated when the current goal is reached and the total number of visited goals is counted. The second task (*defective* case) is like the first, but the sense of direction of one motor is inverted. For each of the three controllers described above ten experimental runs were conducted.

6 Results

The number of visited goals for the two cases (*normal*/*defective*) and the different control paradigms is depicted in figure 3. Sample traces of the robot in the *normal* case are depicted in figure 2. It is in the nature

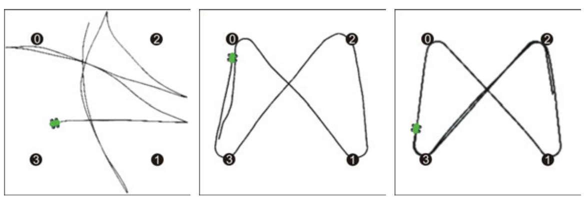

Figure 2: Traces of the robot during an experimental run in *normal* case. The black dots represent the sequence of goals: 0, 1, 2, 3, 0 and so on (robot size scaled up for visibility). Left: pure homeokinetic control (with change of direction after collision), Center: pre- and post-processing framework (see also http://frankhesse.org/jnns), Right: manually defined motor primitives.

of the homeokinetic approach that it is not focusing on the goal (and in this case can not even perceive it) leading to 2.5 visited goals in the mean. The proposed pre- and post-processing framework controlled by H2 clearly shows goal-oriented behavior manifesting in 27 visited goals in the mean. This is slightly lower than with predefined motor primitives (34 goals), due to deviations from the desired path (figure 2, right) by the exploratory drive of the self-organizing control.

In the *defective* case the manually tuned control, relying on predefined primitive behaviors, cannot account for the change in the body properties. Hence the resulting behaviors are not as intended and do not lead to the goals. The proposed pre and post processing framework has no difficulties, since the primitive behaviors can be immediately adapted.

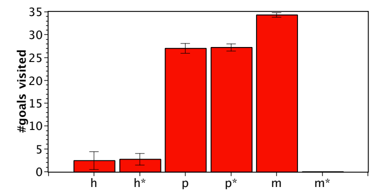

Figure 3: Number of visited goals during ten trials of ten minutes each, h: homeokinetic control, p: pre- and post-processing framework, m: manual defined control. The star indicates experiments in *defective* case, where the sense of direction of one wheel was inverted.

7 Discussion

The presented approach of a neural pre- and post-processing framework provides a possibility to combine self-organization and reinforcement learning for the control of autonomous robots. Instead of predefined motor primitives self-organization based basic motor primitives can be used to achieve a given goal. This requires less initial knowledge about the properties of robot and environment and allows adaptation to changes without relearning of the goal-oriented behavior. So e.g. the system can deal with obstacles/walls or inverted direction of motion of actuators even so they are not addressed in the behavioral control. A combination of self-organization and reward based learning seems thus a promising route for the development of adaptive learning systems.

Note that the proposed framework is not limited to the controllers presented in this study. E.g., recent experiments showed that goal oriented locomotion for a purely central pattern generator driven haxapod robot can be achieved by adding the introduced pre- and post-processing framework.

Acknowledgements

This research was supported by the BFNT Göttingen (project 3B, 01GQ0811), BCCN Göttingen (project D1, 01GQ1005A) and the Emmy Noether Program (DFG, MA4464/3-1).

References

[1] Arkin, R.C. (1998). Behavior-based robotics. *MIT Press, Cambridge*

[2] Der, R. (2001). Self-organized acquisition of situated behaviors. *Theor Biosci, 120(3- 4),179-187.*

[3] Watkins, C. (1989). Learning from delayed rewards. *Phd thesis, Cambridge University, England.*

[4] http://robot.informatik.uni-leipzig.de/software/