Homeokinesis – A new principle to back up evolution with learning

[1]Ralf Der, [2]Ulrich Steinmetz, [2]Frank Pasemann

[1]*University of Leipzig, Institute of Informatics*

[2]*Max-Planck-Institute for Mathematics in the Sciences Leipzig*

Abstract

It is well known that individual learning can speed up artificial evolution enormously. However both supervised learning and reinforcement learning require specific learning goals which usually are not available or difficult to find. We introduce a new principle – homeokinesis – which is completely unspecific and yet induces specific, seemingly goal–oriented behaviors of an agent in a complex external world. The principle is based on the assumption that the agent is equipped with an adaptive model of its behavior. A learning signal for both the model and the controller is derived from the misfit between the real behavior of the agent in the world and that predicted by the model. If the structural complexity of the model is chosen adequately, this misfit is minimized if the agent exhibits a smooth controlled behavior. The principle is explicated by two examples. We moreover discuss how functional modularization emerges in a natural way in a structured system from a mechanism of competition for the best internal representation.

1 Introduction

In recent years, there is a growing interest in using evolutionary strategies for the development of advanced nonlinear controllers. However, between the successful examples treated in the literature and real world applications there exists a complexity barrier which seems insurmountable up to now. By way of example let us consider the case of evolutionary robotics. Despite some interesting achievements, there is also a lot of disappointment concerning the results. So far, all that has been evolved are rather basic behaviors like wall following in mazes, cf. [5, 6, 4]. Notably, the results obtained rest on computer simulations, there is no true *in vivo* evolutionary robotics so far.

Our own practical experience with evolving controllers for the Khepera robot have also revealed the critical role of the fitness function. Usually the latter is designed by hand in view of a certain behavior to be developed by the robot. However quite often the behavior obtained is radically different from expectation. Hence, the design of the fitness function is seldom straightforward and often critical. Co-evolution has recently been discussed as an alternative, cf. [3], but is not general enough to cover typical situations like evolving robots in a *given* hostile environment.

It is well known that learning in the life time of the individuals may accelerate evolution enormously. In fact we think that *in vivo* evolution of robots can not be realized without individual learning. However, usually no teacher is available for the robot and with reinforcement learning one faces the same problem as with the design of the fitness function: Finding the right distribution of rewards is essential for success.

The common feature in the above points is the necessity of explicit goals which have to be formulated carefully in order to drive evolution or learning into the desired direction. In this way one imposes a semantics from outside. Instead we want to find general principles which drive the robot to develop behaviors which so to say make sense in its world without telling the robot what kind of behavior this is to be. Hence the robot discovers a semantics *per se*. Coarsely speaking our idea derives from the observation that with a bad controller the robot (in a maze like situation, say) either gets stuck after a short time or will move more or less chaotically. Both situations mark the limits of the complexity of trajectories the robot can realize. The "good" behaviors of the robot should somehow reside in between these two extremes of complexity.

However there is no hope to use a measure of complexity as an objective function for the learning of the controller. Instead we endow each robot with an *adaptive* model of its own behavior, see Fig. 1. The structure of the model defines implicitly the complexity class of behaviors it can model accurately. Hence we can directly use the model error as a learning signal for the controller, i. e. the controller is adapted such that the behavior of the robot is well "understood" by its model.

We will explicate these ideas in the following. We start with a simple model of linear control of a stochastic dynamical system in order to formulate the above ideas in a transparent manner.

2 Generic example

We consider a mass-less particle with friction under the influence of a harmonic force plus noise

$$\dot{x}(t) = F(x) + \xi(t) \tag{1}$$

the force

$$F(x) = -\kappa_0 x - \kappa_c x = -\kappa x$$

being given by the sum of the forces $F_0(x)$ of the free system and $F_c(x)$ of the controller. $\xi(t)$ is a white Gaussian noise with $\langle \xi(t) \rangle = 0$ and correlation function

$$\langle \xi(t)\, \xi(t') \rangle = \sigma^2 \delta(t - t') \tag{2}$$

where σ^2 is the variance of the stationary distribution of the noise. The free system is stable if $\kappa_0 > 0$, the controlled system being stable if $\kappa > 0$.

Now let us assume that our self-model is chosen such that it can describe the *systematic* behavior of the system (1) accurately. Hence the model does not "understand" the noise. The model can be thought as a predictor evaluating $x_t^{(pred)}(t+\tau)$ which is the predicted value of the observable x at time $t+\tau$ on the basis of the value x observed at time t. The appropriate structural complexity of the self-model is met by using a linear expression $x_t^{(pred)}(t+\tau) = a\, x(t)$. The learning signal (modeling error) is

$$E = \frac{1}{2}\left(x_t^{(pred)}(t+\tau) - x(t+\tau)\right)^2 \tag{3}$$

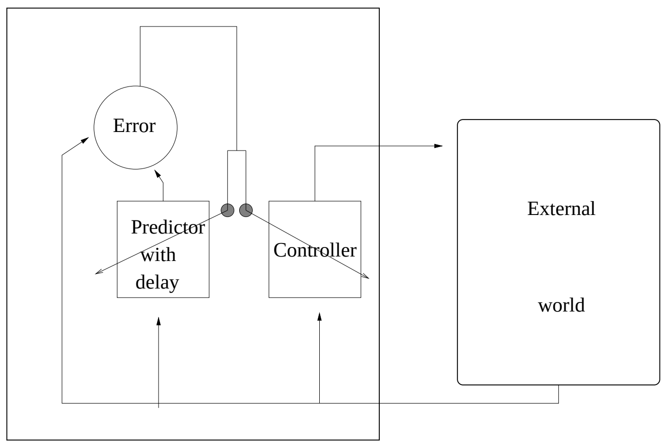

Figure 1: Functional unit with emergent control capabilities. The predictor learns to predict a future state $x_t^{(pred)}(t+1)$ of the external world based on the current observation of the world state $x(t)$. The error unit provides the error signal E according to eq. 3 for the learning of both the predictor and the controller. Behavior (control) is adapted so that the error is minimized. The predictor is of restricted complexity so that it can model only smooth controlled behaviors of the agent in the world. In this way the structural complexity of the predictor defines the kind of behavior that can emerge.

This learning signal is used both for the adaptation of the parameter a and for the adaptation of the controller. We assume that learning of the self-model (predictor) is much faster than that of the controller so that always $a = e^{-\kappa\tau}$.

Averaging over the noise yields

$$E = \sigma^2 \frac{1 - e^{-2\kappa\tau}}{4\kappa\tau} \tag{4}$$

which decays monotonically as a function of κ. Hence learning by gradient descent with respect to κ_c yields a controller which stabilizes the particle at $x = 0$ irrespective of the value of κ_0.

3 The nonlinear case

The emergence of control in the above case may seem an artifact of the simplistic one–dimensional model. However the above analysis has been carried over also to more general cases of linear control with the same result that specific control modes emerge. In order to demonstrate the principle in the nonlinear case we consider the emergence of control with Braitenberg's cybernetic creatures, cf. [1]. We consider the most simple Braitenberg vehicle consisting of two wheels and two sensors. We assume that the right sensor is connected with unit strength with the motor of the left wheel, whereas the left sensor is connected with the right wheel via a coupling of strength w. The light source is moved on the x-axis with velocity chosen such that the distance s to the vehicle which moves with velocity $c = 1$ is kept constant. In this way the angular velocity $\dot{\phi}$ of the vehicle is obtained from the difference of the light intensities s_l and s_r seen by the sensors, s_l being multiplied by the coupling strength w. The equations of motion for the vehicle are

$$\begin{aligned} \dot{x} &= \cos\phi \\ \dot{y} &= \sin\phi \\ \dot{\phi} &= s\left(w s_l\left(y,\phi\right) - s_r\left(y,\phi\right)\right) \end{aligned} \tag{5}$$

where the sensor readings s_l and s_r are

$$s_{l,r} \approx \frac{1}{s^2}\left(1 + \frac{2R}{s}\cos\left(\alpha + \phi \pm \gamma\right)\right)$$

α being the angle between the sensors and $\gamma = \arctan(y/s)$. We consider the case of large s.

For $w = 1$ the vehicle is found to follow the light source in a trajectory which approaches the x-axis in a damped oscillation for all starting values y, ϕ. This is the stable light following behavior as expected by Braitenberg for this case of a symmetric structure of the vehicle with identical couplings. For $w = 1.005$ the vehicle still reaches a stable behavior following the light source however with a transversal offset. For larger asymmetries in the couplings or larger distance form the light source, the vehicle will spiral away from the x-axis, cf. Fig. 2. Obviously Braitenberg's expectation is fulfilled only if the symmetry of the vehicle is nearly perfect. In an evolutionary scenario such beings will exist only if evolution is able of creating perfect individuals. This deadlock is relaxed if the individuals can learn.

Control emerges from our principle already with the most simple predictor proposing $\dot{s}_{l,r} = 0$ for all times. Then the prediction error E simply is the change of the sensor values in one time step τ. Instead of learning a controller we may find the value of w directly from the minimum of E which is a function of both w and the actual coordinates y, ϕ or the sensor values s_l and s_r. This can be done by gradient descent which is achieved by complementing the equations of motion (5) of the vehicle with the differential equation

$$\dot{w} = -\frac{1}{\theta}\frac{\partial}{\partial w} E\left(y, \phi; w\right)$$

the time constant θ being chosen such that the gradient dynamics is fast as compared to the rate of change of vehicle coordinates. As shown in Fig. 2 the adaptation of the behavior in order to satisfy the internal needs of the vehicle (no changes in sensor values) produces a stable light following behavior.

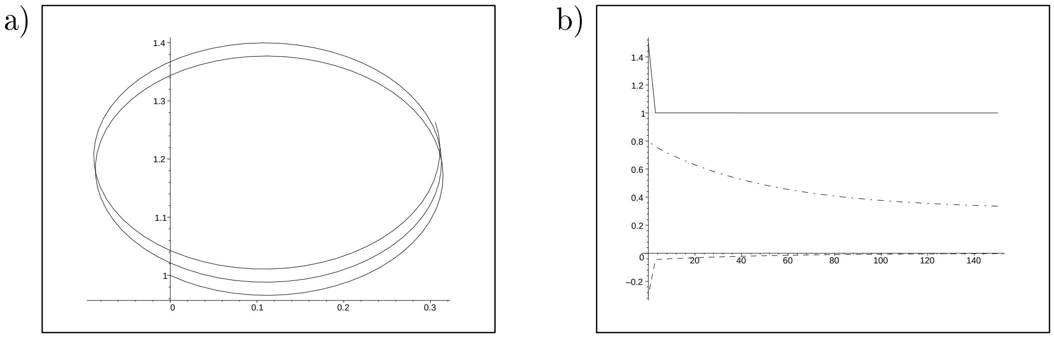

Figure 2: a) Trajectory of a slightly asymmetric Braitenberg vehicle. Obviously the vehicle does not follow the light source which is moving along the x-axis. b) Under control by the homeokinetic principle for a wide range of initial conditions the vehicle develops a stable light following behavior as seen from the behavior of $w(t)$ (top), $y(t)$ (middle) and $\phi(t)$ (bottom).

From the above and several further examples studied we conclude that our principle of using the prediction error as learning signal for the controller is able to produce a variety of potential behavior modes; the structural complexity of the predictor together with environmental conditions decides which mode the controller learns to realize. This effect can be used for the solution of complex nonlinear control problems by task decomposition. The robot or, quite generally, the agent consists of a certain number of functional units of the kind given in Fig. 1 with emergent control capabilities. The principle resembles that of a system of competing experts where each expert develops by itself a certain control behavior. In our approach the structure and number of functional units (experts) are modified as well as the forward/backward characteristics of the predictors. This is done by an evolutionary process which is driven by a very coarse external fitness function.

4 Homeokinesis

In our approach learning is guided by a general principle which is completely unspecific and yet induces specific, seemingly goal–oriented behaviors of an agent in a complex external world. The roots of our principle may be found in the famous principle of homeostasis first introduced by Cannon as early as 1939 [2]. Coarsely speaking, it states that actions of living beings, like food intake or even hunting for food, may be understood in simple terms from an internal perspective: it is driven by the need to keep certain physiological values at a constant level. Deviations from the target values create a control signal which triggers the pertinent actions. So, actions are more or less a by-product of the requirement of keeping internally a stationary state.

Our internal perspective of learning to sustain a smooth controlled behavior in a complex hostile world is that of the adaptive self-model the agent is endowed with. The goal now is *not* to keep the internal system in a stationary state but to learn keeping the agent in a kinetic state (behavior) which is "understood" by the model. In the formulation given in the present paper this principle is a constructive one since it provides a learning signal for the adaptation of both the model and the controller. The learning signal is derived from the misfit between the real behavior of the agent in the world and that predicted by the model. If the structural complexity of the model is chosen adequately, this misfit is minimized if the agent exhibits a smooth controlled behavior.

As in homeostasis, we built up behavior from the internal needs of the agent which here is driven to have a good internal representation of its real world behavior. The difference is that the internal perspective is not rooted in a stationary state but in a kinetic regime. We therefore call our approach the principle of homeokinesis and consider it as the dynamical *pendant* of homeostasis.

5 Concluding remarks

Up to now we have validated our principle both in a variety of linear control tasks by exact analysis like the one given in Sec. 2 and in several nonlinear problems by way of computer simulations. In all cases we observed the emergence of specific control behaviors from our completely unspecific principle. The application to complex nonlinear control tasks as sketched in Sec. 3 is under way. In particular we will present in the near future results on the control of the rotator (generalized pole balancing problem) and the evolution of complex behaviors by autonomous robots.

References

[1] V. Braitenberg. *Vehicles: Experiments in Synthetic Psychology.* MIT Press, 1984.

[2] W. B. Cannon. *The wisdom of the body.* Norton, New York, 1939.

[3] D. Floreano and S. Nolfi. God save the red queen! competition in co-evolutionary robotics. In J. R. Koza, K. Deb, M. Dorigo, D. B. Fogel, M. Garzon, H. Iba, and R. L. Riolo, editors, *Genetic Programming 1997: Proceedings of the Second Annual Conference*, Stanford University, CA, USA, 13-16 July 1997. Morgan Kaufmann.

[4] M. J. Mataric and D. Cliff. Challenges in evolving controllers for physical robots. *Journal of Robotics and Autonomous Systems*, 19(1):67–83, Oct. 1996.

[5] S. Nolfi and D. Parisi. Evolving non-trivial behaviors on real robots: An autonomous robot that picks up objects. *Lecture Notes in Computer Science*, 992:243, 1995.

[6] L. Steels. Emergent functionality in robotic agents through on-line evolution. In R. A. Brooks and P. Maes, editors, *Proceedings of the 4th International Workshop on the Synthesis and Simulation of Living Systems Artificial Life IV*, pages 8–16, Cambridge, MA, USA, July 1994. MIT Press.