Predictive information and explorative behavior of autonomous robots

Nihat Ay[*+], Nils Bertschinger[*], Ralf Der[*#], Frank Güttler[#], Eckehard Olbrich[*]
* Max-Planck Institute for Mathematics in the Sciences Leipzig, Germany
# University Leipzig Institute of Informatics, Germany
+ Santa Fe Institute, Santa Fe, USA

July 26, 2007

Abstract

Measures of complexity are of immediate interest for the field of autonomous robots both as a means to classify the behavior and as an objective function for the autonomous development of robot behavior. In the present paper we consider predictive information as a measure for the behavioral complexity of a two-wheel embodied robot moving in a rectangular arena with several obstacles. The mutual information (MI) between past and future is found empirically to have a maximum for a behavior which is both explorative and sensitive to the environment. This makes predictive information a prospective candidate as objective function for the autonomous development of such behaviors. We derive theoretical expressions for the MI in order to derive an explicit update rule for the gradient ascent dynamics. Theory is based on a model of the sensorimotor dynamics which can be learned on-line. Interestingly in the linear and linearized cases the structure of the learning rule derived depends only on the dynamical properties while the value of the MI enters only as empirical parameter. In this way the problem of the prohibitively large sampling times for information theoretic measures can be circumvented. This result can be generalized and may help to derive explicit learning rules from complexity theoretic measures.

1 Introduction

Complex systems are qualified by a wide variety of structure and/or function. To find convenient measures for quantifying specific properties of such systems is one of the goals of complexity theory. Of particular interest are measures which identify something like the "real" complexity which is small in both very ordered and random systems. There is a long standing tradition linking complexity to information theory [7, 2] and we rely on this parallel in the present paper. Predictive information [1], relating complexity in a time series to the dynamical components that are “meaningful” for prediction, is one prominent example of this paradigm. Technically, the predictive information measures the past-future mutual information in a time series.

These concepts are of immediate interest for the field of autonomous robots. On the one hand we may use complexity theory in order to classify the behavior of robots in interaction with the environment. On the other hand, once such a measure is established it can be used as an objective function for the self-organization of the behavior of the robot. In the present paper we study in some detail the mutual information (MI) between successive time steps which is equal to the predictive information in the case of Markovian systems. The method we are using is to define a robotic system simple enough to be treated analytically but which will be seen to reflect already very much of the full task. In particular our robotic system is fully embodied in the sense that physical influences like inertia, collisions and so on play an essential role.

We consider our contribution as one step towards a systematic approach to the self-organization of the behavior of autonomous robots based on complexity theory. Potential applications of this approach are expected in developmental robotics which has found some interest recently [13] [8]. There is a close relationship to the attempts of guiding autonomous learning by internal reinforcement signals [12] and to task independent learning [10], [11]. Quite generally using a complexity measure as the objective function for the development of a robot corresponds to giving the robot an internal, task independent motivation for the development of its behavior.

The paper is organized as follows: We introduce in Sec. 2 the robot and then give a dynamical systems analysis of its behavior. In particular we introduce the concept of the effective bifurcation point (BP). This analysis is helpful in understanding the different behavioral regimes realized by the robot. Sec. 3 introduces the information theoretic measures and gives a theoretical expression for the case at hand. After this we present in Sec. 4 the results of experiments with the robot showing that the MI has a maximum close to the effective bifurcation point where the robot is seen to cover the largest distances without losing its sensitivity against collisions with the environment. Finally in 5 we formulate a general learning rule for the parameters of the controller based on the gradient ascent of the mutual information as obtained by the theory of Sec. 3. This is seen to be an appropriate way to avoid the sampling problem associated with the empirical MI measure.

2 The robot

In the present paper we are using a simple two-wheel robot simulated by the physics engine ODE. Each wheel is driven by a motor, the motor values being given by the vector $y_t \in \mathbf{R}^2$ which is the output of the controller. The only sensors are wheel counters measuring the true velocity of each of the wheels, i.e. $x_t \in \mathbf{R}^2$ is the vector of the measured rotation velocities. The physics engine simulates in a realistic way effects due to the inertia of the robot, slip and friction effects of the wheels with the ground and the effects of collisions. The velocities are such that the robot upon collisions may tumble so that we have a truly embodied robotic system.

2.1 The control paradigm

There are many different paradigms for the control of autonomous robots. In the present paper we consider closed loop control with a tight sensorimotor coupling. The controller is a function $K : \mathbf{R}^2 \rightarrow \mathbf{R}^2$

$$y_t = K(x_t) \tag{1}$$

where in the present paper

$$K(x) = g(Cx) \tag{2}$$

($g : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ is a vector function, for instance $g_i(u) = \tanh(u_i)$). The matrix C defines the behavior of the robot. In the present paper we use the specific form

$$C = \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix} \tag{3}$$

so that c is the only parameter.

2.2 The sensorimotor loop

In free, unperturbed motion the wheel counters essentially will return the velocities prescribed by the controller, i.e. $x_{t+1} = Ay_t$ where the matrix A is given by $A_{ij} = a\delta_{ij}$ with a a hardware constant which we may set $a = 1$. However in a physical world this will never be the case exactly so that we have to put

$$x_{t+1} = Ay_t + \xi_{t+1} \tag{4}$$

where the matrix A may now be learned by any supervised learning rule and ξ contains all the effects due to friction, slip, inertia and so on which make the response of the robot to its controls uncertain. In particular, if the robot hits an obstacle, the wheels may get totally or partially blocked so that in this case ξ may be large, possible fluctuating with a large amplitude if the wheels are not totally blocked. Moreover ξ will also reveal wether the robot hits a movable or a static object.

Using the above controller the dynamics of the sensorimotor loop can be formulated as

$$x_{t+1} = \psi(x_t) + \xi_{t+1} \tag{5}$$

with ξ introduced in eq. 4 and

$$\psi(x) = Ag(Cx) \tag{6}$$

Although the robot may behave very intricate (see below), eq. 5 is exact, the effects of the embodied interaction with the world being concealed in the model error ξ. In the theoretical analysis given below we will consider ξ as a random number (noise) in order to obtain an explicit expression for the predictive information which forms the basis of our learning rule.

2.3 Properties of the one-dimensional dynamics

Let us now consider at first the extremely simple case of identical wheel velocities, i.e. the robot is moving along a straight line. Dropping the model error (noise) for the moment, we have to find the fixed points (FPs) of eq. (6). With $A_{ij} = a\delta_{ij}$ and $C_{ij} = c\delta_{ij}$, cf. eq. 3, we may consider each loop independently with fixed point equation (put $a = 1$ without restriction of generality)

$$x^* = \tanh(cx^*) \tag{7}$$

Standard FP analysis shows that there is a stable FP $x^* = 0$ for $0 < c < 1$. With $c > 1$ the FP $x^* = 0$ becomes unstable and there are two new, stable FPs $x^* = \pm u$ where for small u we get in leading order by means of the Taylor expansion $\tanh z = z - z^3/3$

$$x^* = \pm\sqrt{3(c-1)}$$

valid for $c = 1 + \zeta$ with $0 < \zeta \ll 1$. On the other hand we also find trivially $x^* \to \pm 1$ for $c \to \infty$.

The behavior of the robot can easily be read off from this analysis. We have the following three scenarios. In the subcritical case, i.e. below the bifurcation point ($c = 1$) the velocity of the robot is fluctuating, due to the noise, around zero with amplitude increasing with c. Hence the robot executes a random walk with variance being the larger the larger c. When encountering a wall it will fluctuate in front of the wall until a longer sequence of random events ξ carries it away.

In the supracritical region with $c \gg 1$ the velocity is fluctuating around one of the stable FPs with amplitude being the smaller the larger c. Hence the robot is moving forever (in physical times) into one direction with more or less constant velocity. Inversion of velocity can take place only if the wheels are totally blocked, i.e. $x_t = 0$ followed by a random event ξ into the appropriate direction. The forces exerted by the robot are very high due to the strong amplification factor c (leading to $y \approx \pm 1$ even if x is already small). Movable objects do not stop the robot so that it can not discern by its behavior between light and heavy movable obstacles.

Eventually, there is a critical region around some value $c_{opt} > 1$ where the noise is able to switch the state between the FPs with a substantial rate. We call this (fuzzy) point the effective bifurcation point. In this region the robot executes long distance sweeps of different lengths into both directions. Due to the smaller amplification rate c, forces are more differentiated so that, by its behavior, the robot may discern between light and heavy movable objects.

It is mainly in the critical region that the robot covers both large distances, into both directions and is sensitive to collisions with an obstacle: If the obstacle is fixed the robot will reverse its velocity (after some time) due to the noise amplification ($c > 1$). If the object is movable the robot will either retract or start moving the object depending on its weight. Due to slip and friction effects, in this critical regime the robot often stops moving the object after some time so that there is a very variate behavior of the robot observed. It is to be noted that these properties, based on proprioceptive sensors (wheel counters) only, are a direct consequence of the closed loop control paradigm used.

2.4 The two-dimensional case

The fixed point analysis obtained for the one-dimensional case readily carries over to the two-wheel robot. Ignoring the noise the controllers of the wheels are completely independent, each controller working only in its sensorimotor loop. Hence with $0 < c < 1$ both sensorimotor loops have FP $z = 0$ and with $c > 1$ we have two FPs for each loop corresponding to the four behavior modes rotating on-site to the left or right and moving forward or backward on a straight line.

With given noise the most interesting regime is observed again about the effective FPs. The robot is expected (and observed) to cover large distances but still reacts sensitively to the collisions with the environment. In particular, by a collision it can be carried over from a straight line to a rotating behavior. The latter can be left if close to the effective bifurcation point. However for large c values, the robot will be caught forever (in physical times) in this rotational mode so that the exploration breaks down. This is well reflected by the results presented in Fig. 3. It is to be noted that, due to physical effects, the two sensorimotor loops are not independent since the wheels are connected by the body. Formally this is contained in the noise ξ. For instance, if the robot collides with some obstacle, the effect on the wheels is strongly correlated. In a head on collision both wheels may be blocked simultaneously which gives a large noise event in both channels simultaneously. Moreover a sudden change in the velocity of one wheel will have an effect on the other wheel due to the inertia effects mediated by the body. This is what makes the present robot more embodied than its one-dimensional counterpart.

3 Information theoretic measures

The aim of the present section is to derive theoretical expressions for the mutual information based on assumptions made on the noise character of the model error of eq. 5. As discussed above, ξ contains the highly nontrivial effects of the embodied robot in interaction with the environment. This does not only mean the presence of higher order statistics but also strong correlations over time (colored noise) due to the inertia of the robot. Nevertheless we assume for the theory a white Gaussian noise. The justification is taken partly from the results. In fact we will see, that with the noise chosen the empirical and theoretical results are in good qualitative agreement. This is sufficient for the present purpose since the theoretical results are used only on the one hand for interpreting the empirical results and on the other hand for the derivation of an on-line learning rule which adapts the parameters of the controller towards the maximum MI regime. Because of the sampling problem this is possible only on the basis of an estimate of the MI with explicit parameter dependence. This (crude) estimate is delivered by our theory.

3.1 The stochastic process in the linear case

Let us first consider again the case of a linear controller, i.e. $g(z) = z$. This is a correct approximation for the case of small z only but will be seen to reveal already much of the nonlinear case. Using the decoupling of the channels and putting without loss of generality $a = 1$, equation (5) reduces for each channel to the first order autoregressive (AR(1)) process

$$x_{t+1} = cx_t + \xi_{t+1} \tag{8}$$

where $|c| < 1$ and we assume that

$$\xi : \mathcal{N}\left(0, \sigma^2\right)$$

Then

$$x : \mathcal{N}\left(0, \frac{\sigma^2}{1-c^2}\right)$$

and

$$p(x) = \frac{1}{\sqrt{2\pi \frac{\sigma^2}{1-c^2}}} \exp\left(-\frac{\left(1-c^2\right)}{2\sigma^2} x^2\right) \tag{9}$$

The conditional probability follows directly from eq. (8)

$$p(x_{t+1}|x_t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_{t+1}-cx_t)^2}{2\sigma^2}} \tag{10}$$

3.2 Predictive information in the linear case

Our system eq. (8) obeys the Markov property so that the full predictive information, cf. [1] relating the future to the past is given by the one-step mutual information

$$\begin{aligned} I(X_{t+1}; X_t) &= \left\langle \log_2 \frac{p(x_{t+1}, x_t)}{p(x_{t+1})\, p(x_t)} \right\rangle = \left\langle \log_2 \frac{p(x_{t+1}|x_t)}{p(x_{t+1})} \right\rangle \\ &= \int\int p(x, s) \log_2 \frac{p(x|s)}{p(x)} dx\; ds \end{aligned} \tag{11}$$

Using Eqs. 9 and 10 we find by elementary means

$$I(X_{t+1}; X_t) = -\frac{1}{2}\log_2\left(1-c^2\right) \tag{12}$$

What does this mean for our robot? Obviously the formula is valid only if our presupposition $c^2 < 1$ is valid. This is clear since the AR(1) process is stationary only if $c^2 < 1$. The MI reaches its maximum if $c^2 = 1$ meaning that the robot amplifies the influence of ξ in a maximum way so that the velocity is fluctuating with maximum possible amplitude. At $c = 1$ the robot executes a random walk with maximum increase in variance under the given noise[1]. Quite generally we may say that in the region $0 < c < 1$ the robots exploration rate increases with increasing MI.

Eq. (12) is correct only if the nonlinearity is sufficiently weak, actually for $0 < c \ll 1$. However the increase of the MI remains true also in the nonlinear case, see the following.

3.3 The nonlinear case

Instead of (8) we consider now the full nonlinear equation (5). While it is not possible to write down a fully analytical solution as in the linear case, we can, however, use the transformation properties of the differential entropy to simplify the expression for the mutual information. If $u = f(v)$ is an invertible function, one has quite generally

$$H(U) = H(V) + \int dv\, p(v)\ \log_2 |J(v)| \tag{13}$$

with $J(v)$ the Jacobian of $f(v)$.
We now have to consider the transformation f from $\{x_t, \xi_{t+1}\}$ to $\{x_t, x_{t+1}\}$, given by (5) and the identity map. The determinant of the Jacobian is 1 and thus the entropy does not change under this transformation:

$$H(X_t, X_{t+1}) = H(X_t, \Xi_{t+1}) \tag{14}$$

If we assume, as in the linear case, that ξ_{t+1} and x_t are statistically independent we get

$$\begin{aligned} H(X_t, X_{t+1}) &= H(X_t) + H(\Xi_{t+1}) \\ H(X_{t+1}|X_t)) &= H(X_t, X_{t+1}) - H(X_t) = H(\Xi_{t+1}) \end{aligned} \tag{15}$$

Now we use that $I(X_{t+1}; X_t) = H(X_{t+1}) - H(X_{t+1}|X_t)$ and get finally

$$I(X_{t+1}; X_t) = H(X_{t+1}) - H(\Xi_{t+1})\ . \tag{16}$$

The entropy $H(X_{t+1})$ has to be evaluated numerically. Fig. 1 shows the result for two different noise strengths. As compared to the empirical behavior of the MI, see Fig. 3, the noise is stronger even in the $\sigma^2 = 0.25$ case so that the maximum is more to the right and is also lower. The agreement is expected to improve with lower noise but the low noise runs are very demanding with respect to time. Nevertheless the results clearly indicate a good qualitative agreement between theory (white Gaussian noise) and embodied experiment as to the position and height of the maximum as a function of the noise strength.

In order to get more explicit theoretical expressions we start again from the linear approximation which is valid only for $|c| \ll 1$ and breaks down completely

[1] We consider only the "physical branch" $c > 0$ in the following.

for $c = 1$ where the MI is infinite. The effect of the nonlinearity for c around 1 is seen by the following argument. We write eq. 5 as

$$x_{t+1} = \tanh(cx_t) + \xi_{t+1} = \gamma(cx_t)\, cx_t + \xi_{t+1}$$

and note that the positive, even function $\gamma(z) = \tanh(z)/z < 1$ so that γ acts as a reduction factor on the value of c which is the smaller the larger x. Approximately we may replace $\gamma(cx)$ with its (time) average so that we get the dynamics equation

$$x_{t+1} = c_{eff} x_t + \xi_{t+1} \tag{17}$$

where $c_{eff} = \overline{\gamma(x)}c$. Obviously $c_{eff} < c$ so that the linear dynamics, eq. 17, can be used as a crude approximation for the full nonlinear dynamics around $c = 1$. This is helpful in order to find an approximation for the MI. In fact, using in eq. (12) c_{eff} instead of c immediately yields an expression for the MI, the quality of which can be studied in Fig. 1 (dashed curve left).

Another approximation is found for large c and small noise by linearizing around the fixed points. Writing $\delta x_t = x_t - x^*$ we get approximately

$$\delta x_{t+1} = L\delta x_t + \xi_{t+1} \tag{18}$$

where

$$L = cg'(cx^*) \tag{19}$$

depends obviously on both $x^*$and c. Let us assume c is far above the bifurcation point and the noise is sufficiently small so that we stay sufficiently long in the vicinity of one of the FPs. Then, using the linearized dynamics and taking account of the bimodality of the distribution we obtain smilarly as in the linear case

$$I(X_{t+1}; X_t) = 1 - \frac{1}{2}\log_2\left(1 - L^2\right) \tag{20}$$

With very large c we may further approximately write $g'(z) = 4e^{-2|z|}$ and $z^* \approx c$ so that approximately

$$L = 4ce^{-2c}$$

and in the same approximation

$$I(X_{t+1}; X_t) = 1 - \frac{1}{2}\log_2\left(1 - L^2\right) \approx 1 + \frac{1}{2\ln 2}L^2 \approx 1 + \frac{8}{\ln 2}c^2 e^{-4c}$$

Obviously, the MI decreases with increasing c exponentially, see the dashed curve in Fig. 1.

4 An embodied robot experiment

It is one of our aims to use the information theoretic measures in realistic robotic applications putting in particular emphasis on the role of the embodiment. This means that we want to discuss physical robots, be it in reality or in simulations,

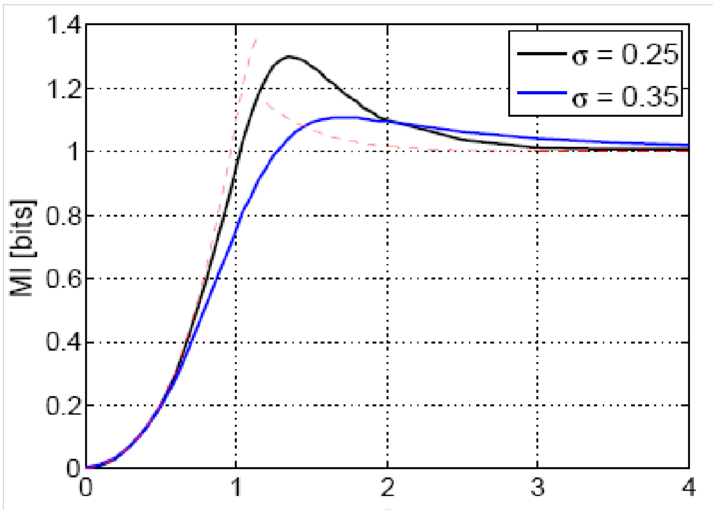

Figure 1: Predictive information of the model sensorimotor loop for two different noise strengths as obtained by computer simulations based on the model dynamics with white Gaussian noise. The maximum, associated with the effective bifurcation point, is seen to move to the right with increasing noise strength. The dashed lines depict the results of the linearized approach. The left hand curve which is approaching the effective bifurcation point from below is obtained with the effective c value as described in the text. The approach from above corresponds to vanishing noise strength.

where the embodiment manifests itself by physical effects like inertia, slip and friction effects, uncertain sensor and actuator functioning. On the other hand we have chosen our experiments such that our theoretical expressions are still useful.

4.1 Experiments

In the experiments, the robot is moving in an arena surrounded by walls and with several obstacles in it so that without any proximity sensors the robot will often collide with either the walls or the obstacles. As discussed in Sec. 2.2, this behavior is largely depending on the value c of the controller (which determines the feed-back strength of the sensorimotor loop).

4.2 The mutual information

A central aim of the present paper was to find the mutual information as a function of c in the embodied robot experiment. In the experiments we evaluated the MI of each of the sensor channels independently. For this purpose we started the robot at a random position and let it run for a long time, mostly for up to

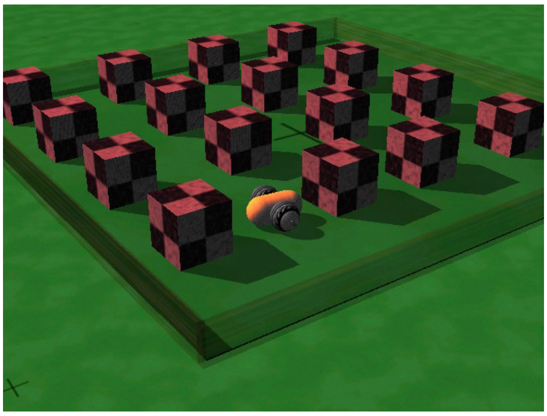

Figure 2: The arena for our two-wheel robot in the starting situation. Obstacles are cubes which are movable upon heavy collisions of the robot. The robot with $c = 1.07$ (maximum mutual information) seems to perform best in the sense of visiting most places in given time (preliminary result) while keeping maximum contact with the environment.

one million steps with a fixed value of c. We discretized the interval of possible sensor values into 30 bins which proved sufficiently accurate by comparison with cases of 10, 20, and 50 bins. Probabilities $p(x)$ or $p(x_{t+1}, x_t)$ were interpreted as relative frequencies of the sensor values in each bin or pair of bins, respectively. The integral in eq. 11 was replaced by the Riemannian sum. The procedure was repeated for every of the c values in the graphics, see Fig. 3.

In practice, the MI was evaluated by an update rule in order to control the convergence progress. Convergence of the MI was reached in typical runs after about 10^5 to 10^6 steps. The convergence largely depends on the value of c. In particular for $c \gg 1$ the robot may change between FPs after a long time and this means that the bimodal distribution may take a very long time until convergence is reached. This problem can be circumvented in a suitable ensemble approach where one may expect that both of the fixed points are populated at the same time.

4.3 Results

The most important result is that in the experiments we find a relatively sharp maximum of the MI at $c_{MI} \approx 1.07$, see Fig. 3. We may interpret the results both from the point of view of an external observer and from the internal perspective of the robot. An external observer might consider different measures which qualify the behavior of the robot. For instance one may ask for the explorative character of the behavior. In our maze like situation we used the average path length covered by the robot in an experiment, which also shows a clear maximum. The maxima of the two curves in Fig. 3 are not exactly at the

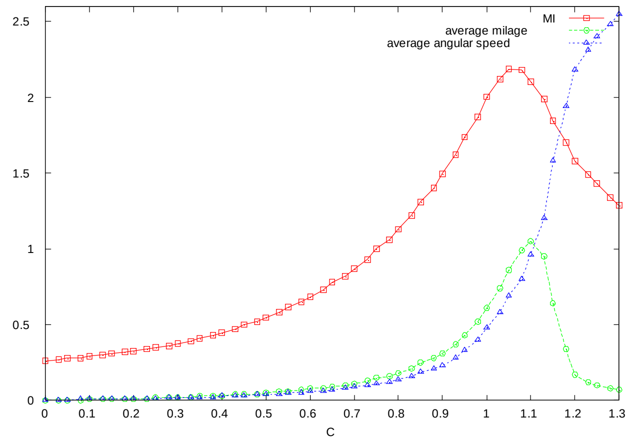

Figure 3: Mutual information and explorativity in an embodied robot experiment: The mutual information (squares) as a function of the parameter c of the controller in the average over both sensors. The maximum of the mutual information is at $c = 1.07$. The strength of the sensor noise was $-0.1 < \xi < 0.1$. The average length of the path covered by the robot (circles) is seen to have a maximum at $c = 1.1$. The average on-site rotation velocity of the robot (triangles) is seen to rapidly increase once c is above the effective bifurcation point.

same position. Instead the MI is maximal at a somewhat lower c value where the robot behaves more cautiously but nevertheless already develops substantial velocities. It is a matter of taste to consider this strategy the better one. In any case we may say that the maximum MI is closely related with the maximum explorativity of the robot. This is our first result.

In future work we will also observe further characteristics of the robot behavior like the distances covered versus the damage probability (overload of the motors, e.g.) and compare those with the predictive information. Alternatively, in cluttered environments, one can also use the frequency by which specific places are visited by the robot. Moreover, the MI we obtained by the simulations is only over one time step. This is identical to the full predictive information in the theoretical model (because of the Markov property) but not in the practical experiments due to the different character of the noise. It is one of the interesting future works to study whether better estimates of the predictive information will reveal new features.

The interpretation from the internal perspective is simple if we use the theoretical results. As shown in Sec. 3.3, according to the simplified model, the variable (depending on c) part of the MI is given by the entropy of the sensor values alone, cf. eq. 16. This means that at the maximum of the MI the behavior is such as to produce the most rich information content of the sensor values. This is our second result which is in nice agreement with other approaches seeing the behavior as a means of structuring input information, cf. Lungarella [9].

Another interpretation in terms of the theory is given by the fact that the maximum of the MI is in the region of the effective bifurcation point. Since the latter is fuzzy we may even identify the two points. The discussion given in Sec. 2 then provides us with a more mechanistic interpretation of the behavior relating maximum mutual information to an effective combination of high sensitivity (to collisions and noise) with high velocities of the robot.

5 Learning rules based on information measures

By our experiments we may conclude that the maximum of the mutual information defines a working regime where the robot is both explorative and sensitive to the environment. This can be used for the construction of a learning rule for the behavioral development of the robot, i.e. we define an update rule for the parameter c as

$$\Delta c = \varepsilon \frac{\partial I\left(X_{t+1}; X_t\right)}{\partial c} \tag{21}$$

We have seen above that the sampling times for the MI are very long so that an on-line learning seems difficult to be realized. On the other hand, when using the theoretical expressions given by eq. 20, we can derive an explicit learning rule for the parameter as

$$\Delta c = \varepsilon_{eff} - 2\varepsilon_{eff} c y_t x_t \tag{22}$$

where $\varepsilon_{eff} > 0$ is an effective learning rate

$$\varepsilon_{eff} = \varepsilon e^{2I(X_{t+1};X_t)} \left(1 - y_t^2\right)$$

Eq. 22 has been obtained by taking the derivative of I only with respect to the explicit c dependence. This can be justified by similar considerations as with the effective c value in Sec. 3.3 but we will not go into the details here. The essential point of this rule is that the MI appears only in the effective learning rate, i.e. it influences only the length but not the direction of the gradient and hence the learning result itself.

The update Δc consists of the driving term ε_{eff} and an anti-Hebbian learning term proportional to $-y_t x_t$. The same learning rule (apart from the prefactor) has been derived in the context of homeokinesis and was discussed in detail elsewhere, cf. [3]. Assuming the system is at the FP, learning is converged if $1 - 2cxg\left(cx\right) = 0$ which, together with $x = g\left(cx\right)$, yields numerically $c = 1.19$. This is a little higher than the empirical maximum of the MI as a result of the approximations made. Nevertheless the learning rule is seen to drive the system close to the working regime where the mutual information is at maximum.

6 Concluding remarks

The aim of the present paper has been twofold. On the one hand we have investigated, in an embodied robot experiment, the role of mutual information between time points as a tool for quantifying the behavior of an autonomous robot. The MI was shown to have a maximum for behaviors where the robot searches its state space in an effective way while still being sufficiently cautious so that collisions with obstacles are not lethal. The latter result is not trivial since, without any bumper sensors, the robot feels the environment only via its wheel counters in a very implicit way mediated by the embodiment.

On the other hand we discussed the complexity measure as the basis for the self-organization of robot behavior by using the measure as an objective function for a gradient following learning rule. The main obstacle in such an attempt are the large sampling times until convergence is reached. In our case we needed 10^5 to 10^6 time steps. Since behavior changes by the learning process, this is prohibitive for any on-line learning scenario. However, our theoretical considerations have shown that, at least in the present case, the structure of the learning rule can be obtained by using a simple model of the sensorimotor loop (which can be learned on-line by any of the known supervised learning procedures) with the mutual information featuring only as some parameter in this rule (here in the effective learning rate). Therefore it seems appropriate to use the crude estimate of the current value of the mutual information given by the theory in order to move, in an on-line learning scenario, towards the maximum of the MI. Once in the region, behavior is changing only slowly so that sampling of the MI will converge partially meaning that the sampled MI may be used for improvements over the estimate.

The generalization of our results to more complicated cases is based on the close relationship of the information theoretic measure to the so called time loop error [4] which has been the basis for concrete learning rules leading to the self-organization of explorative behaviors in complex robots with many degrees of freedom in dynamic, unstructured environments, cf. [6], [5], [3] and the videos on http://robot.informatik.uni-leipzig.de/. We hope in the near future to produce similar results on the basis of information theoretic measures. Preliminary results indicate that the gradients of the time loop error and the mutual information can be related to each other by a change in the metrics of the parameter space.

References

[1] W. Bialek, I. Nemenman, and N. Tishby. Predictability, complexity and learning. *Neural Computation*, 13:2409, 2001.

[2] J. P. Crutchfield and K. Young. Inferring statistical complexity. *Phys. Rev. Lett.*, 63:105–108, 1989.

[3] R. Der, F. Hesse, and G. Martius. Rocking stamper and jumping snake from a dynamical system approach to artificial life. *J. Adaptive Behavior*, 14:105 – 116, 2005.

[4] R. Der and R. Liebscher. True autonomy from self-organized adaptivity. In *Proc. Workshop Biologically Inspired Robotics. The Legacy of W. Grey Walter 14-16 August 2002, HP Bristol Labs*, Bristol, 2002.

[5] R. Der and G. Martius. From motor babbling to purposive actions: Emerging self-exploration in a dynamical systems approach to early robot development. In S. Nolfi, G. Baldassarre, R. Calabretta, J. C. T. Hallam, D. Marocco, J.-A. Meyer, O. Miglino, and D. Parisi, editors, *From Animals to Animats*, volume 4095 of *Lecture Notes in Computer Science*, pages 406–421. Springer, 2006.

[6] R. Der, G. Martius, and F. Hesse. Let it roll – emerging sensorimotor coordination in a spherical robot. In L. M. Rocha, L. S. Yaeger, M. A. Bedau, D. Floreano, R. L. Goldstone, and A. Vespignani, editors, *Artificial Life X : Proceedings of the Tenth International Conference on the Simulation and Synthesis of Living Systems*, pages 192–198. International Society for Artificial Life, MIT Press, August 2006.

[7] P. Grassberger. Toward a quantitative theory of self-generated complexity. *Int. J. Theor. Phys.*, 25(9):907–938, 1986.

[8] M. Lungarella, G. Metta, R. Pfeifer, and G. Sandini. Developmental robotics: a survey. *Connect. Sci.*, 15(4):151–190, 2003.

[9] M. Lungarella, T. Pegors, D. Bulwinkle, and O. Sporns. Methods for quantifying the informational structure of sensory and motor data. *Neuroinformatics*, 3(3):243–262, 2005.

[10] P.-Y. Oudeyer, F. Kaplan, V. V. Hafner, and A. Whyte. The playground experiment: Task-independent development of a curious robot. In D. Bank and L. Meeden, editors, *Proceedings of the AAAI Spring Symposium on Developmental Robotics, 2005, Pages 42-47, Stanford, California, 2005.*, 2005.

[11] J. Schmidhuber. Completely self-referential optimal reinforcement learners. In *ICANN (2)*, pages 223–233, 2005.

[12] A. Stout, G. Konidaris, and A. Barto. Iintrinsically motivated reinforcement learning: A promising framework for developmental robotics. In *The AAAI Spring Symposium on Developmental Robotics*, 2005.

[13] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur, and E. Thelen. Autonomous mental development by robots and animals. *Science*, 291:599 – 600, 2001.