Adaptation of the Perception-Action Loop Using Active Channel Sampling

Philippe Capdepuy[1], Daniel Polani[1,2], Chrystopher L. Nehaniv[1,2]
[1]Adaptive Systems Research Group
[2]Algorithms Research Group
School of Computer Science, University of Hertfordshire
College Lane, Hatfield, Herts, AL10 9AB, UK
{P.Capdepuy,D.Polani,C.L.Nehaniv}@herts.ac.uk

Abstract

During the lifetime of a real world agent or robot, many changes unforeseen at design time can occur. Whether these are due to a change in environmental conditions or to alterations of the embodiment of the robot, flexibility and adaptation are essential qualities that can help it to keep operating in this new situation. This work is based on an information-theoretic approach and introduces an exploration strategy that allows an agent to detect and adapt to changes in its perception-action loop by actively sampling areas of interest. We define the problem of exploring the sensorimotor channel and establish a measure of the distance between the observed and the real model of the channel. An optimal Oracle-based strategy is used to compare performances of the adaptive sampling strategy and a random strategy. Results for different scenarios of change in a binary channel show that the proposed strategy is highly effective in many cases. We also outline principles to adapt this mechanism to the exploration of multiple channels and we give preliminary results for such a scenario.

1 Introduction

The development of adaptive sensorics (and actuatorics) is a topic of high current interest and relevance. The advent of increasing powerful and ubiquitous computational resources has brought about the ability to construct hardware of many different sizes for a variety of use niches. This makes it increasingly important to provide this growing number of individual (and interconnected) devices with the ability to interact flexibly and adaptively. At this point, most of the activities in this direction have to be explicitly engineered: any adaptivity of a device has been planted into it by the manufacturers, any flexibility of reaction requires a protocol that specifies how a device is to handle novel stimuli and unforeseen situations. "True" adaptivity, in the sense of a device "learning on its own" is still very much elusive; existing device adaptivity relies on engineered failure/success models of devices.

In this dilemma, inspiration from biology is sought: biology has a seemingly unmatched reservoir of successful adaptation strategies. Evolution is probably the most celebrated of these, but there are many more: whether Neural Networks, Ant Colony Optimization, Artificial Immune Systems, or other paradigms, there is a rich variety of methodologies that have originally been motivated from the biological example.

While these paradigms share the general biological motivation, they have, structurally, little in common and it seems difficult to formulate a common principle which gives rise to them. This implies that any even bio-inspired adaptive algorithm used in an engineering problem needs to be hand-fitted to the problem at hand.

However, in the last years evidence has been mounting that even the convoluted dynamics of biological adaptation may be governed by simple fundamental principles; even more interestingly, some of these principles are well established in engineering, namely as principles of (Shannon) information optimization. For instance information maximization principles (infomax) give rise to biologically plausible neural receptive fields [16], or neural codes [18, 4, 7, 3, 21]. The latter seem to operate at the trade-off curve between information transmission and metabolic cost [15] and, more than that, organisms are ready to trade off a very significant amount of information (in typical cases of

978-0-7695-3166-3/08 $25.00 © 2008 IEEE
DOI 10.1109/AHS.2008.64

the order of magnitude of 10%-20% of the organism's total metabolic energy) to acquire sensoric and process it [14]. This indicates that (Shannon) information is a vital resource for organisms, almost on par with its metabolic energy. Why should that be the case? The main hypothesis is that of a principle of *parsimony*: of two organisms which e.g. utilize the same amount of metabolic energy it is likely that the organism which makes better use of the available information will have an evolutionary advantage. In absence of any evolutionary advantage of that information, the metabolic cost of processing the given information can be deevolved by degenerating the associated neural and sensoric apparatus (as happens with cave fish).

Such a parsimony principle provides a way of understanding what needs to happen in an adaptive system that mimics biological operation. However, there is another interesting factor involved: the influence of the environment on the organism does not reflect the standard view of a sender and receiver communicating with each other using a common code [8]. Rather environment and organism/agent interact in a quite intricate manner which nevertheless can be captured by novel mathematical formalisms: the treatment of information processing in the perception-action loop of agents can be modeled transparently by the use of causal Bayesian Networks [11, 10] which extend Ashby's Law of Requisite Variety [1, 22, 23] to general sensorimotor loops. This provides a handle for a quantitative treatment of general infomax scenarios of an agent and thus an approach towards a systematical, but yet biologically relevant methodology for constructing adaptive devices.

As a particularly promising path, the use of *empowerment* has been suggested [12, 13, 5], a concept similar to the channel capacity of the external part of the perception-action loop of an agent (we discuss this formally and in detail in Sec. 2). Empowerment measures by how much (in terms of information) an agent can *potentially* modify its environment so that it is able to register this modification. Essentially, empowerment quantifies a combined controllability/observability [17, 19] in information-theoretical terms.

Empowerment has been shown in a range of scenarios to constitute a universal utility that, if maximized locally, provides behaviours consistent with the natural choice of humans in a "self-motivated" way (not unlike the homeokinetic principle [6], autotelic principle [20] or the learning progress maximization [9] or the predictive information [2]). The reason for the success of empowerment is not fully understood at this time, although first hypotheses are emerging.

The current paper, however, will not preoccupy itself with this question — it will assume that, as evidence seems to indicate, the central hypothesis is valid that empowerment is indeed a quantity of interest to induce adaptive behaviour in an agent embedded in an environment via its sensorimotor loop. Up to now, all earlier scenarios studied calculated empowerment separately or externally. Once done, they assumed that, for the duration of a particular behaviour strategy, the empowerment profile of the system would stay unchanged. Real systems will be different — the reaction of the environment to the actions of an agent (even if in the same states) may change with time. In such cases, all the relevant quantities of the perception-action loop need to be reestimated for empowerment to be up-to-date. The current paper will discuss how to adaptively and efficiently estimate the relevant signatures of a perception-action loop. Section 2 introduces the information-theoretic perspective of the perception-action loop. In Sec. 3 we define the exploration problem and introduce a measure of the performance of exploration. The optimal Oracle-based policy and the adaptive exploration strategy are then introduced. The performance of the latter is then evaluated in different scenarios against the optimal strategy and a random one. Section 4 describes an adaptation of the exploration problem to multiple channels related through a topology of contexts and shows some preliminary results for a simple grid world.

2 The Information-Theoretic Picture of the Perception-Action Loop

We will refer to the perception-action loop of the agent as a causal Bayesian network which describes the relationships between the environment, the sensors and the actuators of the agent. The perception-action loop can then be unrolled in time (see Fig. 1) and some of its properties can be assessed using information-theoretic tools. One central aspect of our work is to investigate the sensorimotor channel, i.e. the channel that goes from actions to future perceptions through the environment. An important characterization of this channel is provided by the concept of empowerment [12, 13]. The idea is to measure how much information can be injected by an agent into its environment and then perceived back through its sensors. More precisely it is defined as the channel capacity from the sequence of actions $A_t, A_{t+1}, \ldots, A_{t+n-1}$ to the perceptions S_{t+n} after a fixed number of time steps. The channel capacity is defined as the maximum mutual information between the sent message and the received message, where the maximization is made with respect to the probabilities

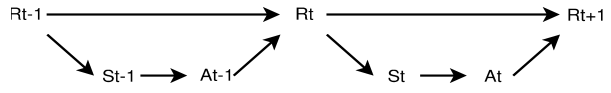

Figure 1. Representation of the perception-action loop as a causal Bayesian network unrolled in time. R_t stands for the environment of the system, S_t is the sensor of the agent and A_t its actuator.

for the sent message. In the context of this work, we will restrict ourselves to the simplest case where only the current action and the next sensoric state matter. Empowerment can then be written as

$$\mathfrak{E}(A_t \rightarrow S_{t+1}) = \sup_{p(a_t)} I(A_t; S_{t+1}) \tag{1}$$

with $p(a_t)$ the probability distribution function of the action. Empowerment can be described as the maximum potential information an agent can transfer into its own sensors through the environment.

In the perception-action loop, the properties of the channel that goes from actions to future perceptions depend on both the embodiment of the agent and its coupling with the environment. In the case of a real agent these properties, described as the conditional probability distribution $p(s_{t+1}|a_t)$, are subject to changes due to alterations of the embodiment or changes in the environment. If only observational data are available and if the channel is unstable, estimating empowerment becomes a difficult task. To get good estimates of empowerment, an accurate model of the environment is necessary. The purpose of this work is to provide an active exploration strategy that maximizes the accuracy of the constructed model.

3 Exploration as Sampling of the Perception-Action Loop

In all this section we will use a conceptually simple case, the single channel case, to define the basic principles of our exploration strategy. The perspective taken in this work is to consider an agent that constructs a statistical model of its perception-action loop by collecting samples. This model is represented by a probability distribution $p(s|a)$ (precisely it is $p(s_{t+1}|a_t)$ but for the sake of clarity we will use the short version) with $s \in \mathcal{S}$ being the perceptual space, and $a \in \mathcal{A}$ the set of possible actions. To construct this model, the agent has to explore the channel by acting on it. At each time-step it picks an action and sends it into the channel, through the environment, and then perceives back a particular sensor value.

By collecting such data it is possible to approximate the real probability distribution of the channel (if it is stationary). However, if one supposes that the channel can sometimes be changed (e.g. external damage, change in the environment) then the agent has to reevaluate its statistical model to reflect the changes and match the new real model. We make the assumption that the channel is changed to another almost stationary channel.

In the following subsections, we formalize what are the real and the observed model and define a measure of their distance. Using this measure we can establish an Oracle-based optimal strategy for exploration. Subsequently we propose a simple heuristic that allows to approximate this strategy. Efficiency of this heuristic is then evaluated against the optimal strategy and a purely random one.

3.1 Real and Observed World

The whole point of an exploration strategy when used on its own is to provide the explorer with an accurate model of its environment. Basically the world can be described as a model, and the subjective vision of the explorer is another model. The purpose of exploration is to minimize the distance between the real and the observed model. In the single channel case, the real world model is represented by a probability distribution $p_r(s|a)$ and the agent model is constructed by sampling the channel, leading to another probability distribution $p_o(s|a)$.

As our goal is to maximize the accuracy of the observed channel, we need a way of measuring how much the two models match. For this purpose we use the Jensen-Shannon distance between the two distributions, averaged over all actions (which we will consider equiprobable). The Jensen-Shannon distance is based on the Kullback-Leibler divergence between two distributions p and q, defined by

$$D_{KL}(P||Q) = \sum_i p(i) \frac{\log p(i)}{\log q(i)}, \tag{2}$$

but where the distance is the average of the divergence between each distribution and their average M:

$$D_{JS}(P||Q) = \frac{1}{2}\Big(D_{KL}(P||M) + D_{KL}(Q||M)\Big). \tag{3}$$

where $M = \frac{1}{2}(P+Q)$. We can therefore measure the distance ϵ between the real and the observed model

using

$$\epsilon(P_o||P_r) = \frac{1}{|\mathcal{A}|}\sum_a D_{JS}\big(P_o(S|a)\big|\big|P_r(S|a)\big). \quad (4)$$

3.2 Defining an Optimal Sampling Strategy

Now that the problem has been stated and that we have a measure of the distance between the observed and the real model, we can define what we will consider as an optimal strategy. The goal of the exploration strategy is to match as quickly as possible the real world model by sampling it with actions. An optimal strategy is one that would maximally reduce this distance at each sampling.

If one considers that there exists an Oracle who knows the real model of the environment, one can define a strategy that will use this Oracle to pick the actions which are more likely to have an informative outcome (in the sense that it will change our current knowledge). Formally we define the change in accuracy δ_ϵ when performing action a and observing outcome s (i.e. by adding a new sample at time t) by

$$\delta_\epsilon(a,s) = \epsilon\Big(P_o|S_{t+1}=s, A_t=a \quad \Big|\Big| \quad P_r\Big) - \epsilon(P_o||P_r) \quad (5)$$

where $P_o|S_{t+1}=s, A_t=a$ is the observed model after being updated with the new sample. According to the real model of the environment we can define for each action a the expectation of change in accuracy by

$$E\big[\delta_\epsilon|A_t=a\big] = \sum_s p_r(s|a)\delta_\epsilon(a,s). \quad (6)$$

As our goal is to minimize the distance between the observed and the real model, the optimal Oracle-based strategy is to pick the action a which has minimum $E[\delta_\epsilon|A_t=a]$. Of course a real agent does not have access to the Oracle, however this strategy will be useful in our case to evaluate the performance of other strategies.

3.3 Approximating the Optimal Sampling Strategy

Now comes the central question. How can an agent that has no access to an Oracle discover an efficient sampling strategy. The goal of the agent is also to minimize the distance between his observed model and the real one, but it has no access to this distance measure. One way to obtain information that is relevant to this problem is to consider not only the current observed model, but also how it evolves in time.

In the case of an agent that has a model that perfectly matches the environment, the sampling process will not bring anything new, i.e. it will not change the model (apart from small fluctuations, but this problem is addressed at the end of the paper). However if the model of the agent is not accurate for a particular action, sampling this action will provoke strong changes in the distribution of sensoric outcomes. By taking into account this time evolution, the agent can estimate how accurate the different parts of its model are, and therefore have an idea about the ϵ function that only the Oracle detains. From the agent perspective we can make the following assumption: if a part of our model changed due to recent sampling, then our model was (and probably still is) not accurate. Therefore if we want to maximize the accuracy of our model, this part needs more sampling in order to converge to the real distribution.

The key idea of our approach is to quantify these changes in the distribution and then to use this quantity as a guide to pick the action that is most likely to get us to the real model. To measure the change in the probability distribution, we use the variation of the entropy of the distribution. Formally, for a given action a, and an observed outcome s, the entropy variation of the corresponding distribution is

$$\delta_H(a,s) = H(P_o|S_{t+1}=s, A_t=a) - H(P_o|a) \quad (7)$$

where H stands for the Shannon entropy

$$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log p(x). \quad (8)$$

To use this heuristic, the agent simply has to favor actions which are changing the model, i.e. actions for which $|\delta_H|$ is maximum. The absolute value is taken because we do not care if the entropy is increasing or decreasing, what we care about is if it changes at all.

3.4 Results for the Single Channel Case

To evaluate our heuristic (δ_H), we compare it with a random strategy (which always converges after a sufficiently long time) and the Oracle-based strategy. The experiment consists in providing initial data from a particular channel, assuming that it is known perfectly by the agent, and then changing the channel and letting the agent explore it. We measure how much time each strategy takes to converge to 1% of the initial error ϵ.

We use a collection of different binary channels described in table 1 that have different properties in term of randomness. For each pair of different channels (one used as initial channel and the other used as the

Table 1. Binary channels used for evaluation, separated in deterministic channels, half deterministic, and completely random.

Name	$p(S=1\|A=0)$	$p(S=1\|A=1)$
ID	0	1
NOT	1	0
ZERO	0	0
ONE	1	1
HID0	0	$\frac{1}{2}$
HID1	$\frac{1}{2}$	1
HNOT0	1	$\frac{1}{2}$
HNOT1	$\frac{1}{2}$	0
RAND	$\frac{1}{2}$	$\frac{1}{2}$

changed channel) we perform 100 experiments and average the measures. We use the Oracle-based strategy as a baseline for the speed of convergence, and we express the result for the random strategy and the δ_H as the ratio between their convergence time and the baseline. The δ_H strategy is in fact an ϵ-greedy strategy with $\epsilon = 0.1$, meaning that 90% of the time the agent picks the action that has maximum $|\delta_H|$ and a random action the rest of the time. Results are described in tables 2 and 3.

For every combination of channels studied, the δ_H strategy clearly outperforms the random strategy. On average the δ_H strategy takes 9% more time than the baseline Oracle-based strategy, whereas the random strategy takes on average 62% more time. Qualitatively it is possible to classify the different scenarios into two main groups. The first group includes all the channel changes that involve a modification of the outcomes for both actions. In this group the random and the δ_H strategy have close results, but the δ_H strategy still outperforms the random one, having an average ratio of 1.05 against 1.23. But the real effectiveness of the δ_H strategy appears when changes are only partial (in this case when only one of the actions has a different outcome after the channel change). In this case it has an average ratio of 1.13 against 2.02 for the random strategy.

If one action has been changed but the other stayed the same, then only for the first one will the entropy change and therefore it will be sampled more often. In the case where both actions are changed we obtain a slightly more complex behaviour. This is the case for the scenario ID to HNOT0 (see Fig. 2). In this scenario the outcome of both actions are changed. For action 0 the outcome changes from a deterministic (only 0) to the opposite deterministic distribution (only 1). On the other hand, action 1 changes from a deterministic outcome (1) to a random one. We can observe on the graph that the behaviour of the δ_H strategy differs quite a lot from the Oracle-based and the random ones. The two latter strategies sample both actions at very similar frequencies whereas the δ_H strategy strongly changes over time. To understand it better we now describe in detail what is happening (we suggest the reader to first have a look at Fig. 3 to have a graphical representation of the problem).

At the beginning, both distributions have first to move from a deterministic low-entropy distribution to a high-entropy random one. However as the outcome of action 0 is always 1, it moves faster toward the maximum entropy state than action 1 does, leading to higher δ_H. Therefore during the first 40 time-steps of simulation sampling is dominated by action 0. When this distribution gets close to the maximum entropy one, its derivate diminishes, making action 1 the most sampled during the next 300 hundreds time-steps. At

Table 2. Ratio between the convergence time of ($Random$;δ_h) and the baseline time provided by the Oracle-based strategy for each scenario. Rows represent the initial channel, columns correspond to the channel after the change. The second part of the results is shown in table 3.

	ID	NOT	ZERO	ONE	HID0
ID	—	1.0;1.0	2.0;1.1	2.0;1.1	2.0;1.1
NOT	1.0;1.0	—	2.0;1.1	2.0;1.1	1.5;1.1
ZERO	2.0;1.1	2.0;1.1	—	1.0;1.0	2.0;1.0
ONE	2.0;1.1	2.0;1.1	1.0;1.0	—	1.5;1.1
HID0	2.0;1.1	1.0;1.0	2.0;1.1	1.0;1.0	—
HID1	2.0;1.1	1.0;1.0	1.0;1.0	2.0;1.1	1.5;1.1
HNOT0	1.0;1.0	2.0;1.1	1.0;1.0	2.0;1.1	2.0;1.2
HNOT1	1.0;1.0	2.0;1.1	2.0;1.1	1.0;1.0	1.5;1.1
RAND	1.0;1.0	1.0;1.0	1.0;1.0	1.0;1.0	2.0;1.4

Table 3. Continuation of table 2.

	HID1	HNOT0	HNOT1	RAND
ID	2.0;1.0	1.5;1.1	1.5;1.1	1.1;1.0
NOT	1.5;1.1	1.8;1.0	2.2;1.0	1.1;1.0
ZERO	1.5;1.1	1.4;1.1	1.8;1.0	1.0;1.0
ONE	2.0;1.0	2.1;1.1	1.5;1.1	1.1;1.0
HID0	1.6;1.1	2.0;1.6	1.5;1.1	1.9;1.3
HID1	—	1.5;1.2	2.0;1.2	2.1;1.2
HNOT0	1.5;1.1	—	1.5;1.1	2.2;1.3
HNOT1	2.0;1.2	1.6;1.1	—	2.3;1.4
RAND	2.0;1.4	2.0;1.3	2.0;1.4	—

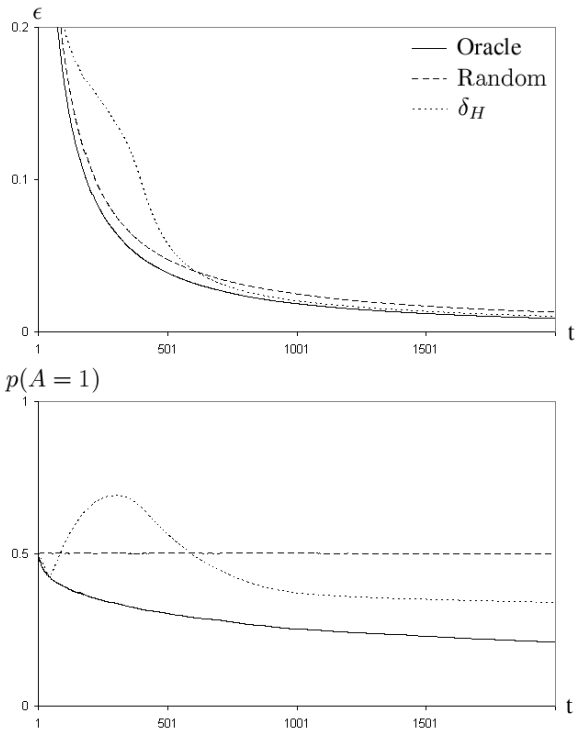

Figure 2. Scenario with a complete change of the channel (ID to HNOT0) averaged over 100 experiments (2000 time-steps). Top: time evolution of distance between observed and real model for the different strategies. During the first 600 time-steps, the δ_H-strategy deviates from the two others. After this time it overtakes the random strategy and gets close to the optimal one. Bottom: proportion of actions sampled (see text and Fig. 3 for details)

.

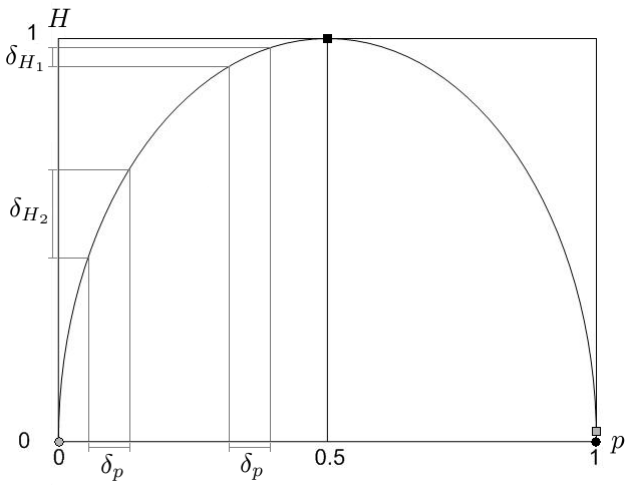

Figure 3. Entropy function of a binary distribution (using base two logarithm). For a given probability change δ_p due to a new sample, the entropy change δ_H depends on the previous state of the observed distribution. When entropy is maximum (i.e. $p = 0.5$), δ_H reaches a minimum. Dots represent the distribution for action 0 (circle) and 1 (square) of the initial (gray) and changed (black) channels in the scenario ID to HNOT0. During the sampling process, the gray dots converge toward their black couterparts.

this point, action 1 is getting very close to the maximum entropy level where it has to converge. However sampling of action 0 starts to move from the maximum entropy (and low derivative) toward the lower entropy value where it converges. By doing so the derivative grows leading to a positive feedback effect that reinforces exploration of action 0. Eventually both distributions converge toward the changed channel.

4 Multiple Channels

Now we consider a more complex case: exploration of multiple channels. By multiple channels we do not mean that the agent has a number of constantly accessible channels for which it has to get a model, in that case we would simply consider them as one composite channel and use exactly the same strategy as for the single channel case. In this section we are interested in situations where channels are not all directly accessible to the agent but instead it has to move between channels by performing actions (and sampling at the same time). For such a case we will refer to the concept of *contexts*. We assume that the agent is able to distinguish different contexts (for example based on the current sensoric state) and that each context c is associated with a particular channel.

We first define how the contexts are related to each other through a topology and we translate the problem of channel exploration to this topology. Two cases are distinguished, the first one is the general case where the channels and their topology are not related. The second one is a particular case where the channels and their corresponding topology are completely intertwined. This case has very important connections with models of the perception-action loop and empowerment maximization. We then introduce a simple mechanism to use the δ_H-strategy in such topologies. Again, simu-

lation results for simple scenarios are used to compare the different strategies.

4.1 Context Topology

We introduce a principle which we refer to as *context topology*. The idea is the following, for the sampling agent the world is represented as a collection of separate channels $c \in \mathcal{C}$ similar to the ones described in the previous section but uniquely identified by a context. When the agent is in a particular context, it performs an action to sample the corresponding channel. The difference with the single channel case is that the action will not only bring a new perceptive sample but it might also move the agent in a different channel. The context topology is described by the probability distribution $p(c_{t+1}|a_t, c_t)$ and it can also be subject to changes.

4.2 Propagating the Sampling Strategy

The goal of the agent is still to maximize the accuracy of its model $p(s_{t+1}|a_t)$ where a_t is an action and s_{t+1} is the sensor state obtained after performing the action. However now there are multiple channels and for all of them we have to maximize the accuracy. To adapt the δ_H-strategy to this topology of channels, we use a framework similar to that of reinforcement learning. For each channel-action pair we associate a 'reward' value which is simply the last entropy change of the distribution associated with this action in this context $\delta_H(a, c)$. This value is then propagated into the topology by using a value-iteration algorithm:

foreach $c \in \mathcal{C}$ **do**
  $V(c) \leftarrow 0$;
end
repeat
  $\Delta \leftarrow 0$;
  foreach $c \in \mathcal{C}$ **do**
    $V'(c) \leftarrow max_a\Big(\delta_H(a, c) + \gamma \sum_{c_{t+1}} p(c_{t+1}|c_t, a_t)V(c)\Big)$;
    $\Delta \leftarrow max(\Delta, |V'(c) - V(c)|)$;
  end
  $V = V'$;
until $\Delta < \theta$;

Algorithm 1: Value iteration algorithm in the multichannel case.

In this algorithm γ is the discount factor and θ is a small number that stops the algorithm when a sufficient precision has been reached. When the agent is in context c, the action-selection process consists in picking the action a that maximizes the utility quantity

$$U(a, c) = \delta_H(a, c) + \gamma \sum_{c_{t+1}} p(c_{t+1}|c_t, a_t)V(c). \quad (9)$$

4.3 Preliminary results

We evaluate this model in a simple grid world with a moving agent. The agent senses its absolute position in the world and it can move to any neighboring cell (if not occupied by a block) or stay in the same cell. The current sensor value is used as the context. Initially the grid world is surrounded by blocks, preventing the agent to move out of it, but the inside is empty. We allow the agent to collect statistics about this initial environment. After some time we introduce a block inside the box, changing the channels that are located next to this block.

The experimental setup consists of a 11 by 11 grid world and we performed 100 experiments during which we measured the distance between the observed and the real model during 1000 time-steps. To avoid being stuck sampling areas already very close to the real value, we used a Boltzmann selection instead of the ϵ-greedy strategy. In a given context c the probability of picking action a is defined as $p(a) = \frac{1}{Z}e^{U(a)/T}$ where Z is a normalization factor $Z = \sum_{a'} e^{U(a')/T}$, T is a temperature parameter, and $U(a, c)$ is the utility calculated by the value-iteration algorithm.

Parameter values used for this experiment are $T = 0.01$, $\gamma = 0.8$ and $\theta = 0.001$. We measured the distance between the observed and the real model for the δ_H strategy and the random strategy at the end of the experiment. Values obtained for the δ_H strategy are significantly better, 24% of the initial distance, than the random strategy that reaches on average 58% of the initial distance. These preliminary results are encouraging but a more systematic study is needed to properly assess the effectiveness of the δ_H strategy in such multichannel case.

5 Conclusion

In the context of agents constructing a model of their perception-action loop by collecting statistics, we have proposed an active sampling strategy (δ_H) based on the temporal change of the entropy of the model. This strategy allows an agent to quickly adapt to changes of their perception-action loop. As the perception-action loop of the agent reflects its embodiment and the coupling with the environment, any change in the environment or any damage to the sensoric of actuatoric

apparatus of the agent can impact the model of the perception-action loop. Using the proposed adaptive sampling strategy, the agent will reinforce exploration of these changes in order to quickly converge to the new model.

We first performed a set of experiments on different scenarios of change with a single binary channel case and measured the convergence time for the different strategies. The results for the δ_H strategy are very close to the optimal Oracle-based strategy (9% more time); comparatively, the random strategy performed quite poorly (62% more time). The behaviour of this strategy has been detailed in some particular scenarios.

We extended the δ_H strategy to the exploration of multiple channels related to each other by a context topology. Preliminary results on a simple grid world show that the proposed strategy performs significantly better than a random one. However more results are needed to validate its efficiency in different scenarios.

Future investigations will focus on the use of such an exploration strategy for maximization of empowerment. We expect this model to extend results in the area of self-organization in collective systems (as has been investigated in [5]). Useful applications of this model also include sensor evolution scenarios, where different sensorimotor apparatus can be evaluated in a given environment and compared on such criteria as stability of perception-action loop model and potential capacity to inject information in future sensoric states.

References

[1] W. R. Ashby. *An Introduction to Cybernetics*. Chapman & Hall Ltd., 1956.

[2] N. Ay, N. Bertschinger, R. Der, F. Güttler, and E. Olbrich. Predictive information and explorative behavior of autonomous robots. *European Journal of Physics B*, 2008. Submitted.

[3] W. Bialek, R. R. de Ruyter van Steveninck, and N. Tishby. Efficient representation as a design principle for neural coding and computation. arXiv.org:0712.4381 [q-bio.NC], December 2007.

[4] N. Brenner, W. Bialek, and R. de Ruyter van Steveninck. Adaptive rescaling optimizes information transmission. *Neuron*, 26:695–702, 2000.

[5] P. Capdepuy, D. Polani, and C. Nehaniv. Maximization of potential information flow as a universal utility for collective behaviour. In *2007 IEEE Symposium on Artificial Life*. IEEE, 2007.

[6] R. Der. Self-organized acquisition of situated behavior. *Theory Biosci.*, 120:1–9, 2001.

[7] A. L. Fairhall, G. D. Lewen, W. Bialek, and R. de Ruyter van Steveninck. Efficiency and ambiguity in an adaptive neural code. *Nature*, 412:787–792, 2001.

[8] J. J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin Company, Boston, 1979.

[9] F. Kaplan and P.-Y. Oudeyer. Maximizing learning progress: an internal reward system for development. In F. Iida, R. Pfeifer, L. Steels, and Y. Kuniyoshi, editors, *Embodied Artificial Intelligence*, volume 3139 of *LNAI*, pages 259–270. Springer, 2004.

[10] A. Klyubin, D. Polani, and C. Nehaniv. Representations of space and time in the maximization of information flow in the perception-action loop. *Neural Computation*, 19(9):2387–2432, 2007.

[11] A. S. Klyubin, D. Polani, and C. L. Nehaniv. Organization of the information flow in the perception-action loop of evolved agents. In *Proceedings of 2004 NASA/DoD Conference on Evolvable Hardware*, pages 177–180. IEEE Computer Society, 2004.

[12] A. S. Klyubin, D. Polani, and C. L. Nehaniv. 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.

[13] A. S. Klyubin, D. Polani, and C. L. Nehaniv. 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.

[14] S. B. Laughlin. Energy as a constraint on the coding and processing of sensory information. *Current Opinion in Neurobiology*, 11:475–480, 2001.

[15] S. B. Laughlin, R. R. de Ruyter van Steveninck, and J. C. Anderson. The metabolic cost of neural information. *Nature Neuroscience*, 1(1):36–41, 1998.

[16] R. Linsker. Self-organization in a perceptual network. *Computer*, 21(3):105–117, March 1988.

[17] A. I. Mees. *Dynamics of feedback systems*. John Wiley & sons, Ltd., New York, 1981.

[18] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek. *Spikes*. A Bradford Book. MIT Press, 1999.

[19] E. D. Sontag. *Mathematical control theory; Deterministic finite dimensional systems (Texts in Applied Mathematics)*, volume 6. Springer-Verlag, New York, 1990.

[20] L. Steels. The autotelic principle. In F. Iida, R. Pfeifer, L. Steels, and Y. Kuniyoshi, editors, *Embodied Artificial Intelligence: Dagstuhl Castle, Germany, July 7-11, 2003*, volume 3139 of *Lecture Notes in AI*, pages 231–242. Springer Verlag, Berlin, 2004.

[21] S. F. Taylor, N. Tishby, and W. Bialek. Information and fitness. arXiv.org:0712.4382 [q-bio.PE], December 2007.

[22] H. Touchette and S. Lloyd. Information-theoretic limits of control. *Phys. Rev. Lett.*, 84:1156, 2000.

[23] H. Touchette and S. Lloyd. Information-theoretic approach to the study of control systems. *Physica A*, 331:140–172, 2004.