Improvement Proposals to Intrinsically Motivationnal Robotics

Fabien Hervouet, Eric Bourreau

To cite this version:

Improvement Proposals to Intrinsically Motivationnal Robotics

Fabien Hervouet and Eric Bourreau
LIRMM – Université Montpellier 2
`fabien.hervouet@lirmm.fr, eric.bourreau@lirmm.fr`

Abstract—In this paper, we first give a brief background on artificial curiosity and intrinsic motivation as it is studied in the developmental robotics research community. We then introduce some theoretical improvements for mechanisms related to intrinsically motivational living algorithms as introduced by SAGG-RIAC algorithm. Finally we conclude by drawing the way for our future experimental results using these new improvements.

Index Terms: developmental robotics, autonomy, intrinsic motivation, curiosity

I. INTRODUCTION

The goal of any developmental robotics algorithm is to design a control loop that – if executed on a physical robot – should reveal its own body affordance within its environment, through an obviously unsupervised process. We also believe that the key challenge is to identify and implement low-level mechanisms that allow a long-term development. This means we have to inspect and design scalable task-independent mechanims that may involve the robot in a self autonomous skill practice driving it through its whole life.

In the following section, we give a brief background on artificial curiosity and intrinsic motivation focused on the approach originally proposed by *Oudeyer* and the *IAC* algorithm [3]. Then we introduce new ideas like intrinsically diversificative interest measure or new way of computing action for goal reaching. We also introduce the idea of memory restructuring, with a symmetric dynamic measure for both splitting and merging sensory regions. We then conclude by drawing the way for our future experimentations using these new improvements.

II. IAC AND SAGG-RIAC

Soon after the birth of the developmental robotics community [4], Oudeyer and Kaplan introduced the *IAC* (*Intelligent Adaptive Curiosity*) algorithm [3]. The mechanism is anchored at a sensorimotor level and allows low-level action selection in the high-dimensional sensorimotor space for a robot. This algorithm postulates that one way to provide autonomy to a robot is to let it make its proper action choices, based on its experience, in order to maximize its learning. This approach provides a fresh and applicative perspective in the sensori-motor learning for robots. With this architecture an embodied agent is going to experiment sensorimotor coordinations in order to learn the effects of its actions thanks to a unique action selection mechanism that tends to choose actions that improve prediction quality. It means that an unexperienced agent will only have to pursue the goal of executing actions driven by its *structural coupling history* with the environment and an *interest measure*. At the beginning there is only one region that represents the global sensorimotor space, i.e. the union of the sensor domains and actuator domains. Step-by-step this mother region is going to be split into subregions. This process can be considered as a specialization phenomenon. Indeed, the agent initially has a coarse and inaccurate view of its sensorimotor space, and through experience, it will tend to divide the space in order to evaluate the impact of certain parameters in executing actions.

Baranes and Oudeyer then proposed a few years after an evolution to the original *IAC* algorithm called *SAGG-RIAC* [1] which is of interest to us because of the competence acquisition paradigm it explores. The global principle remains the same except that this time, the agent has to choose sensory regions where it wants to return to. Practically, the *SAGG-RIAC* algorithm is based on alternating reaching phases (i.e. reaching a goal in what they call an operational space) and local exploration phases (i.e. improving the world comprehension toward the goal). The purpose of reaching phases is to test the reliability of the forward motor model while the purpose of exploration phases is to improve the inverse model of the system. Exploration phases are triggered when the reliability of the local controller is too low. In the following section we explain some improvements to this algorithm we would like to introduce and experiment.

III. IMPROVEMENTS TO SAGG-RIAC

Although we keep the overall operation of the motivationnal living algorithm SAGG-RIAC [1], we draw some improvements. Thus we introduce a new measure of interest directly integrating a diversification mechanism, we propose a new simple way of computing an action driving toward goal reaching, and finally we explain our process of memory restructuring with splitting and merging mechanisms.

A. Interest measure

The measure of interest ρ qualifies the dynamic interest of a region. Our measure is based on the one introduced by Baranes and Oudeyer in [1] with two major differences.

$$\rho(R_i) = learningProgress(R_i) + diversification(R_i)$$

First we propose to use a timestamped derivative. This tends to reduce the interest by flattening the interest curve when

experiments are very infrequent. Let c_j and t_j respectively the competence and the timestamp of the jth experiment stored in memory, with $|R_i|$ the number of experiments in a region.

$$learningProgress(R_i) = \frac{\sum_{j=0}^{|R_i|/2} c_j - \sum_{j=|R_i|/2}^{|R_i|} c_j}{\sum_{j=0}^{|R_i|/2} t_j - \sum_{j=|R_i|/2}^{|R_i|} t_j}$$

We also chose to incorporate a UCT based [2] diversification measure which takes into account in an incremental way the number of experiments conducted in the current region relative to the total number of experiments, instead of choosing a region goal with external probabilistic diversification as it is done with *SAGG-RIAC*. Let n and n_i be respectively the total number of experiments and the number of experiments in the current region. Let c be a constant that allows to adjust the result of diversification measure.

$$diversification(R_i) = c \times \sqrt{\frac{\ln n}{n_i}}$$

B. Getting next action to reach a goal

We propose to compute action towards a goal using k-nearest-neighbour experiments chosen among previously acquired – from explorations phases – experiments. These experiments must maximize two criteria: the initial and final states should be as close as possible respectively to current and goal states. The strategy then consists in generating a mean action with respect to actions performed in these filtered experiments.

C. Memory organization and restructuring

A typical developmental and incremental process pushes the robot to start its learning from scratch. This implies that the robot acts very strangely at the begining because it does not hold enough information about the world it develops in. Therefore, pushing the robot to split its sensorimotor space allows it to overcome the lack of information at the beginning of the developmental living process. This is this particular splitting condition that makes the strength of this approach because it allows *the isolation of coherent experiments*[1]. But if a region is necessarily split as soon as it contains a certain amount of experiments as done by Baranes and Oudeyer, it means that this process does not take into account whether the region is already sufficiently small to maximize the coherence. Whatever knowledge and experience accumulated, the agent will always tend to split the space pretexting there is nothing to learn about how to (re)organize the memory through the developmental process.

Therefore two mechanisms come to mind to prevent from over-splitting: (1) *improve the splitting condition so as to make it dynamic*, i.e. correlated with the development of the agent, and (2) *propose a mechanism for merging regions to allow a subsequent restructuring*. We chose to combine these two mechanisms by proposing to introduce a joint measure for both splitting and merging:

$$\mu(R_1, R_2, R) = \frac{|LP(R_1) - LP(R_2)|}{LP(R)}$$

---

[1]Coherence depends on the nature of the measure that determines the best split in a sensorimotor region. Here it is related to the notion of learning progress, i.e. the derivative of learning.

This measure means that in the splitting case we will split the current region R if it contains two subregions R_1 and R_2 exhibiting better learning progress, relatively to the current learning progress in the region R. On the other hand in the merging case, we will merge two regions that have been splitted possibly a long time ago when the agent didn't hold enough information: the current region R_1 should be merged with another region R_2 into a sole region R if the learning progress of R exhibits a better evolution than R_1 and R_2.

In order to be able to easily implement both of these mechanisms we chose to introduce a graph structure of sensorimotor regions that upgrades the tree structure of sensorimotor regions used since *IAC*. We believe that, by its more scalable and flexible nature, the graph structure facilitates from an algorithmic point of view the merging of non-convex and even non-adjacent sensory regions.

IV. CONCLUSIONS & PERSPECTIVES

In this paper we argue that in the developmental robotics community, and more specifically in the sub-community interested in intrinsically motivated robotics, there is a way to propose some improvements. Thus we proposed to integrate a diversification mechanism directly in the interest measure. We proposed a new simple way to compute the next action for reaching a goal based on previously acquired experience. Finally we proposed a joint measure to handle discovering and forgetting. Discovering process is underlined by a better dynamic splitting condition that fits developmental view. Forgetting does not mean that the agent should forget any experience information acquired during its interaction in the environment. We rather believe that the agent should forget parts of the segmentation of the sensorimotor space that used to make sense with a lack of information but that seems inappropriate as it is more experienced. For this purpose we proposed to upgrade typical tree structure memory with a graph structure.

Even if we haven't been able yet to conduct experiments providing sufficient interesting data to be analyzed, this is exactly what we are willing to do in the near future. We want to reimplement and compare the different ideas proposed in the community with our own ones. We will also work at grasping the operational and behavioral gap between simulated and physical robotic experiments. Finally, the next big challenge will be to *define, evaluate and compare performance* of several agents' developmental trajectory.

REFERENCES

[1] BARANES, A., AND OUDEYER, P.-Y. Intrinsically-motivated goal exploration for active motor learning in robots: A case study. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010)* (2010).

[2] KOCSIS, L., AND SZEPESVÁRI, C. Bandit based monte-carlo planning. In *In: ECML-06. Number 4212 in LNCS* (2006), Springer, pp. 282–293.

[3] OUDEYER, P.-Y., AND KAPLAN, F. Intelligent adaptive curiosity: a source of self-development. In *Proceedings of the 4th International Workshop on Epigenetic Robotics* (2004), L. Berthouze, H. Kozima, C. G. Prince, G. Sandini, G. Stojanov, G. Metta, and C. Balkenius, Eds., vol. 117, Lund University Cognitive Studies, pp. 127–130.

[4] WENG, J., MCCLELLAND, J., PENTLAND, A., SPORNS, O., STOCKMAN, I., SUR, M., AND THELEN, E. Autonomous mental development by robots and animals. *Science 291* (2001).