Intrisically Motivated Goal Space Creation for Autonomous Goal-Directed Exploration in High-Dimensional Unbounded Sensorimotor Spaces

Fabien Benureau
ENS Lyon / INRIA, France
fabien.benureau@inria.fr

Pierre-Yves Oudeyer
INRIA, France
pierre-yves.oudeyer@inria.fr

I. INTRODUCTION

A. Motivation

Today's robotic systems are given increasingly complex tasks in an increasing variety of situations such as object or social interaction. Many of those situations cannot be anticipated at design time : autonomous learning capacities are needed to adapt to novel, unexpected conditions. Yet, because of their complex bodies and multiple sensors, robots face highly-dimensional, unbounded, continuous sensorimotor spaces whose semantics are often unknown.

Such spaces are too large to be explored exhaustively, an issue even more crucial in robotics given the expensive and slow nature of the physical interactions needed to gather training data. Learning in those spaces also raises other challenges, because robot's sensorimotors spaces are highly heterogeneous and multi-modal, with unreachable areas because of physical constraints, unlearnable areas because the actions of the agent do not have any influence on the sensors values, and yet other area where learning is made difficult by huge noise-to-signal ratios or requires the previous aquisition of other skills (e.g. learning reaching before grasping).

This is why efficient explorations techniques are needed, where each interaction maximize the knowledge or competence gained through each interaction. To adress this issue, statistical learning techniques have focused on optimizing exploration policies to maximize various criteria in particular through *active learning* [1]-[3].

Another approach have stemmed from the field of developmental robotics, where inspiration from psychology and neuroscience research on animal and infant learning [4] [5] [6] have highlighted the importance of curiosity in skill acquisition. Several intrinsically motivated learning techniques have been proposed [10] [11] [12].

In this article, we will build on a particular intrinsically motivated, goal-oriented technique initiated by Baranes and Oudeyer [7], which defines the interest of an area of the sensorimotor space as the progress of the competence in reaching self-assigned goals in this area. This method has yielded excellent results in experiment with motor spaces of high dimension. Yet sensory spaces have remained limited to 2 or 3 dimensions, and the robot had only one type of action to consider. Moreover, the goal space was predefined by hand.

We propose a broad expansion of the previous architecture, where the sensory space has 10+ dimensions, and relevant goal space are created and their interest evaluated by the algorithms through novel techniques. Additionally, we considers robotic agents that have several different actions at their disposal that can combine them temporally. To our knowledge, no existing work addresses both those challenges.

B. Problem

We consider a robot equiped with multiple sensors and several motors. The robot has several different action at its disposal, henceforth called *motor primitives*, which are hardcoded sequences of motor orders whose behavior is controlled through continuous parameters. Motor primitives have been shown to exist in biological organisms [8]. For example, in a wheeled robot, a motor primitive could control how and how long a specific wheel spins, with a target velocity and a duration parameter.

A motor primitive has a specific number of parameters with fixed boundary values. An motor primitive order is described by the motor primitive, a value for each parameter, and a start date. The start date delays the execution of the motor primitive. For example, given the primitive *Spin(d, v)* controlling a wheel with d the duration of the primitive, and v the target velocity of the wheel, a fully qualified order is *(Spin, (3.5, 1.0), 2.0)*, which instruct the motor to make start making the wheel spin at 1.0 *rad/s* at time 2 *s*, and for 3.5 *s*. Every motor primitive is designed to end after a finite amount of time.

The robot is able to combine several motor primitives by sending multiple orders at once. With the start date of each primitive, the robot can temporally stage the parallel execution of a set of orders. This temporal combination is simple, yet expressive and effective. Even with a small number of primtives, and given the continuous nature of the parameters, the space of possible orders is extremely large.

We are interested in the robot's ability to learn how to reach any reachable state of the sensors, given access to its previous observations. Since, given unlimited ressources, a brute force method testing every order combination at a reasonable resolution would achieve that, we are interested in the efficiency of the process.

II. Algorithms

The learning process consist of repeated interactions. The interest mechanism chooses a goal, i.e. a set of target sensors values. The learning algorithm then produce a set of orders aimed at achieving that goal. The orders are executed, and the effect of the orders - the final state of the sensors - is paired to the orders, creating an observation. The learning and interest internal data is then updated with the new observation. The robot is then reset in initial conditions and another interaction can take place.

A. Learning

Our learning algorithm is based on local linear assumptions allowing global non-linear modeling. Once a goal has been set, the observation with the nearest effect to the goal effect is searched, and a neighborhood of the orders corresponding to the nearest effect is created amongst past observations. Using standard linear regression and constrained optimization techniques [9] (since orders parameter legal values are bounded), we then compute a set of orders aimed at reaching the goal.

B. Curiosity

Our curiosity architecture proceedes from two main ideas.

The first one is that high-dimensional spaces necessitate too many observations to obtain useful learning data. Indeed our learning algorithm requires set of local observations in order to make local linear assumptions. In a high-dimensional space, such a local neighborhood only happens when an enormous amount of data has been collected. Instead we create small subspaces whose dimensions are a subset of the dimensions of the global sensory space. This creation is autonomous. For a given global sensory space, we obtain a collection of small subspaces from which to draw goals. Baranes and Oudeyer [7] have defined an interest measure on the areas of a goal space. We extend this measure to goal spaces themselves, so the probability of drawing a goal from a particular area of a particular goal space is expressed by :

$$P(\text{goal space})P(\text{goal area}|\text{goal space})$$

where each probability is proportional to the interest of the space or area concerned.

Then, after orders have been executed, at the end of an iteration, when, we have and orders set/effect pair, we can update every subspace with a new observation since we read every sensors, i.e. every dimension, for each observation.

The second one is that we need to quickly create areas of competence. Even those low-dimension subspaces are unbounded, and sampling them with a good enough resolution is probably impossible at the timescale of the learning process. In order to become competent in those subspaces, we create small areas where observations are concentrated. Then, we draw goals around these areas in order to make those area of competence grow. This allow the robot to be competent in at least in some areas at any moment during the learning process: this competence can be used as the basis of a scaffolding learning processus, or as useful data for a planning mechanism. Finally, it helps delimiting reachable areas from unreachable ones. It also help quickly differentiate regions, instead of keeping a fairly homogeneous and low interest measure across the whole space. Interesting or uninteresting regions are quickly identified as such, since goals get drawn to a smaller region of the whole space.

III. Experimental Set-up

To test our ideas, we consider a dual-wheeled robot, equiped with an arm that can throw a ball. The robot is given three motor primitives, one for each wheel with velocity and duration as parameters, and one to throw the ball with angle and force as parameters. With every parameters and the start date of each motor primitive, we have a motor space of dimension 9.

The sensor of the robot consist of its position (x, y), its orientation θ, the speed of each wheel motor (v_x, v_y), the value of two accelerometers placed on the main chassis near each wheel (a^L_x, a^L_y) and (a^R_x, a^R_y) and the position of the ball, relatively to the position of the robot (b_x, b_y). The sensory space is thus of dimension 11.

In order to quickly test our ideas, we use a simple simulation of such a robot using a kinematic model. At the time of writing, we did not collect sufficient experimental data to produce results with a good measure of confidence.

References

[1] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, *Active learning with statistical models*, J. Artif. Intell. Res., vol. 4, pp. 129145, 1996.

[2] B. Settles, *Active Learning Literature Survey*, Univ. Wisconsin-Madison, Madison, WI, 2009, CS Tech. Rep. 1648.

[3] R. Castro and R. Novak, *Minimax bounds for active learning*, IEEE Trans. Inf. Theory, vol. 54, no. 5, May 2008.

[4] R. W. White, *Motivation reconsidered: The concept of competence*, Psychol. Rev., vol. 66, pp. 297333, 1959.

[5] J. Lehman and K. Stanley, *Abandoning objectives: Evolution through the search for novelty alone*, Evol. Comput. J., 2010, to be published.

[6] P. Redgrave and K. Gurney, *The short-latency dopamine signal: A role in discovering novel actions?*, Nature Rev. Neurosci., vol. 7, pp. 967975.

[7] A. Baranes and P.-Y. Oudeyer, *R-IAC: Robust Intrinsically motivated exploration and active learning*, IEEE Trans. Autonom. Mental Develop., vol. 1, no. 3, pp. 155169, 2009.

[8] J. Kongzak, *On the notion of motor primitives in humans and robots* in Fifth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems, Lund University Cognitive Studies, vol. 123, pp. 47-53, 2005.

[9] M.J.D. Powell (1994), *A direct search optimization method that models the objective and constraint functions by linear interpolation*, in Advances in Optimization and Numerical Analysis, eds. S. Gomez and J-P. Hennart, Kluwer Academic (Dordrecht), pp. 5167

[10] J. Schmidhuber, *Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts*, Connect. Sci., vol. 18, no. 2, pp. 173187, Jun. 2006.

[11] S. Singh, A. Barto, and N. Chentanez, *Intrinsically motivated reinforcement learning*, in Proc. 18th Annu. Conf. Neural Info. Process. Syst., Vancouver, BC, Canada, 2004.

[12] P.-Y. Oudeyer, F. Kaplan, and V. Hafner, *Intrinsic motivation systems for autonomous mental development*, IEEE Trans. Evol. Comput., vol. 11, no. 2, pp. 265286, Apr. 2007.