Towards hierarchical curiosity-driven exploration of sensorimotor models

Sébastien Forestier, Pierre-Yves Oudeyer

Towards Hierarchical Curiosity-Driven Exploration of Sensorimotor Models

Sébastien Forestier, Pierre-Yves Oudeyer

Flowers Team — INRIA Bordeaux

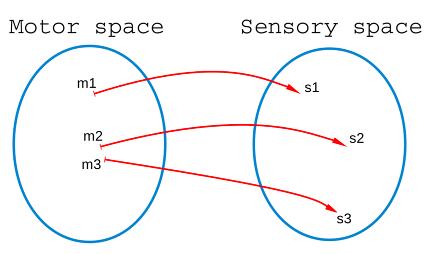
Introduction

Curiosity-driven exploration mechanisms have been proposed to allow robots to **actively** explore high dimensional sensorimotor spaces in an open-ended manner [1]. In such setups, **intrinsic motivations** based on **competence progress** show good results - the learner explores its sensory space with a bias toward regions which are predicted to yield a high competence progress. However, throughout its life, a developmental robot has to **incrementally explore skills** that add up to the hierarchy of previously learned skills, with a constraint being the cost of experimentation. We rely on the SAGG-RIAC series of architectures [2] and describe some ways to **extend** those architectures to the exploration of a hierarchy of sensorimotor skills. We developed a **simulated robotic setup** to evaluate the different architectures, where a robot has to push an object to different locations.

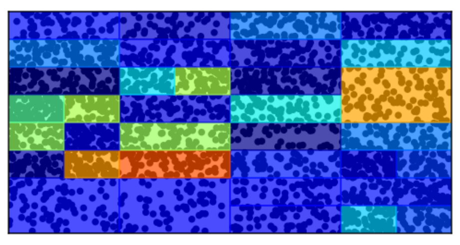
Curiosity-Driven Exploration

We use and extend the Explauto library [3] that aims at studying autonomous exploration. In the Explauto framework, a sensorimotor model is learned together with an interest model that guides future exploration.

- **Sensorimotor model**
  This model stores the experimented motor commands and their associated sensory experience and builds a mapping between the motor space and the sensory space. We use the fast nearest neighbor algorithm to build this mapping but more powerful regression methods could be used instead.



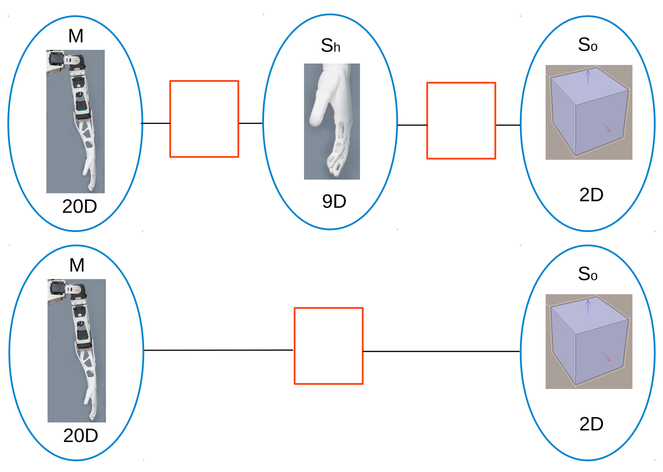
- **Interest model**
  The interest model estimates how interesting it is to explore given parts of the sensory space. We use the SAGG-RIAC architecture (Self Adaptive Goal Generation - Robust Intelligent Active Curiosity [2]) with an intrinsic motivation that pushes the agent to explore regions where the progress of the competence to reach self-generated goals in the sensory space is the higher.



- **Exploration**
  The agent selects a goal in its sensory space according to the interest model and infers motor parameters to get close to this point with the sensorimotor model. It adds some exploration noise to discover new motor configurations, executes the command and observes the sensory experience. Finally, it updates the sensorimotor model with the new association, and the interest model with the competence to reach the goal.

Hierarchical Exploration

To study hierarchical exploration, we design a setup where the **agent learns two sensorimotor models**, with the second one that reuse the first one. In our setup, the first sensorimotor model in the hierarchy is a relation between the 20 motor parameters of a robotic arm and the 9 parameters of the $3D$ trajectory of the robotic hand. The second sensorimotor model is a relation between the trajectory of the hand and the $2D$ position of a block at the end of the movement. Control architectures will have to learn directly a relation between the motor parameters and the position of the block.

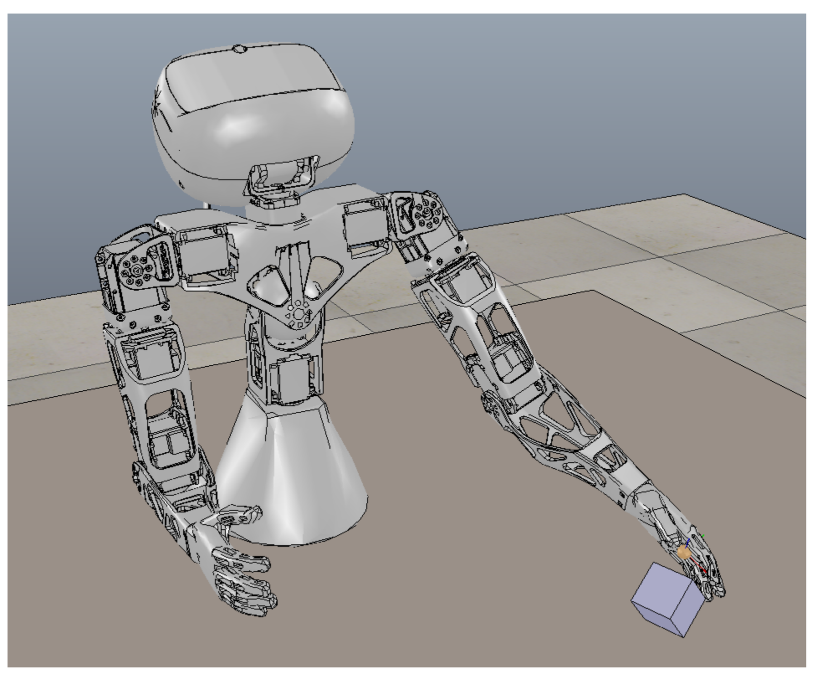


Learning Architectures

- **Motor Babbling Control**
  Exploration of a mapping between M and S_o with random motor actions.
- **Goal Babbling Control**
  Exploration of a mapping between M and S_o with a competence-based intrinsic motivation (SAGG-RIAC).
- **Simplest First**
  Exploration of a mapping between M and S_h for the first half of the trials and of a mapping between S_h and S_o for the second half. Both explorations are made with SAGG-RIAC.
- **Top-Down Guidance**
  Exploration of a mapping between S_h and S_o. When asking to explore a hand's movement s_h, this goal s_h is used to guide the exploration of a mapping between M and S_h which is given a certain amount of iterations to try motor configurations to reach that goal (with a black-box optimization technique [4]).

Poppy Torso in the V-rep simulator

- The robotic setup is the left **arm of the Poppy robot** [5], with 4 degrees of freedom, simulated with the V-REP simulator based on the Bullet physics engine.
- We use Dynamical Movement Primitive [6] to **control** the arm's movement as this framework allows the production of a diversity of arm's trajectories with few parameters. Each arm's motor is controlled by a DMP with a starting and a goal position equal to the rest position of the motor. Each DMP is parameterized by one weight on each of 5 basis functions whose centers are distributed homogeneously throughout the movement of duration $4s$. M is the 20D space of the motor parameters.



- S_h is a 9D space representing the 3D trajectory of the **hand**. We also use the DMP framework to project each of the X, Y and Z movements on a sensory DMP with 3 basis functions.
- A $3cm$ wide **block** is located near the robot's hand and can be moved in different complex ways, e.g. with the hand pushing on the top of the block or on a side. S_o is the 2D space representing the position of the block at the end of the simulation. The initial position of the block is $(X = 0.225m, Y = 0.135m)$, X axis points in front of the robot, Y axis on its left.

Experiments

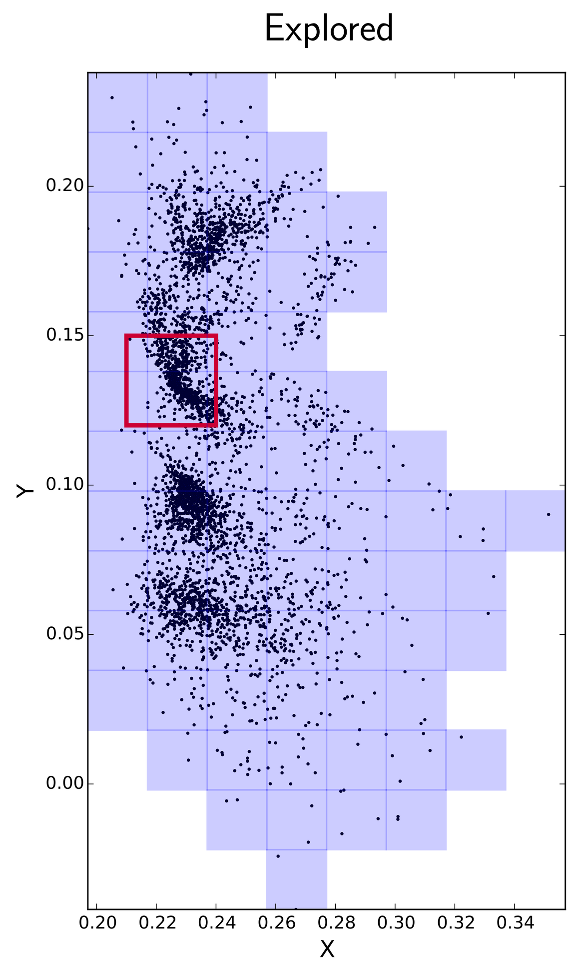
We ran different trials for each conditions ($n_{MB} = 13$, $n_{GB} = 9$, $n_{SF} = 16$, $n_{TDG} = 18$). Each trial is made of 5000 learning iterations. Each iteration takes about $1s$ on $3,06\ GHz$ Xeon x5675 nodes. Exploration is evaluated each 500 iterations with 2 different measures. Statistical tests are performed after 5000 iterations to compare conditions.

Measures

- **Exploration Measure**
  This measures how **diverse** outcomes have been found by the agent in its goal space. It makes no assumptions about the interest of exploring different regions of the goal space from the viewpoint of the engineer. To compute the quantity of exploration we consider the goal space as an unbounded grid where we count the cells that have been reached during training.

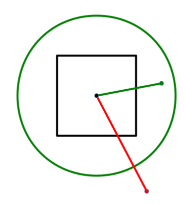
  Low Robustness
  In our setup, the **environment is highly stochastic** in the sense that when the agent tries the same motor command different times, the variability in the generated arm movement and the collisions between the arm and the block leads to high variances in the end position of the block. Thus the agent might not succeed in reaching again a previously reached position, and in other words, a high exploration does not mean a high learning. We define the following measure to tackle this problem.

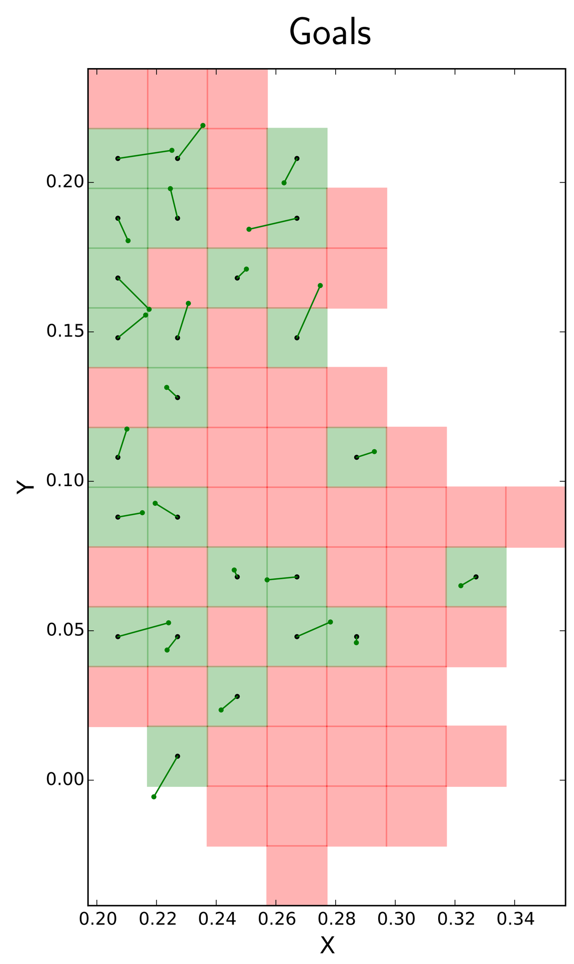
- **Competent Exploration Measure**
  This measures how competent is the robot on the explored part of its goal space. We give the centers (black dots) of explored cells as goals to the agent, and we count the cells where the agent manages to get close to the goal (within $2cm$).



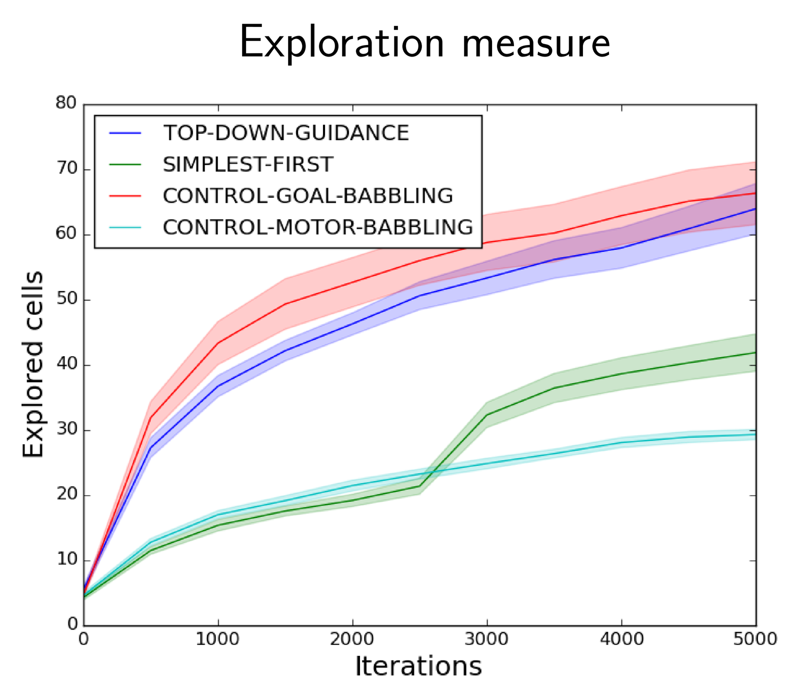
Left: Red square is the initial position of the block, blue dots are reached positions of the block at the end of the movement, blue squares are reached cells.

Right: In the competent exploration measure, we plot a red cell if the agent did not manage to get within $2cm$ of the center, and a green cell if it did. In green cells, black dots are goals and green dots are corresponding reached points. X axis points in front of the robot, Y axis on its left, units are meters.

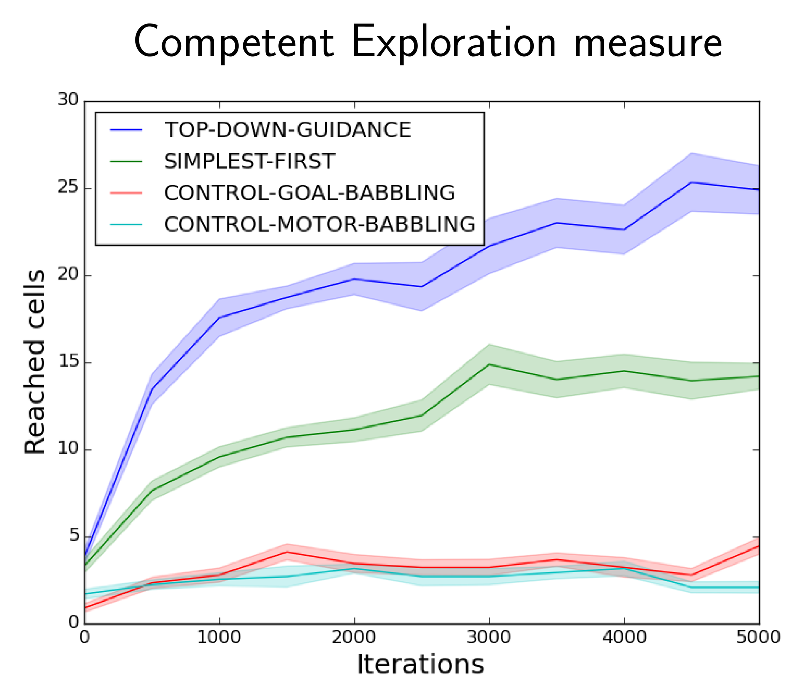




Results



Mann-Whitney U tests at the end of the experiments show that the exploration measures in the Goal Babbling control and Top-Down Guidance conditions are not significantly different but are significantly greater ($p < 0.05$) than in the Simplest First condition, where it is significantly greater than in the Motor Babbling control condition.



Mann-Whitney U tests at the end of the experiments show that the competent exploration measure is significantly greater in the Top-Down Guidance condition than in the Simplest First condition where it is significantly greater than in the control conditions.

Discussion

- Results show that exploration is better in the Top-Down Guidance condition than in the Simplest First condition, with both measures. This implies that the **guidance from the top-level model is useful to drive the exploration of the lower-level model** on parts of the task space that are more interesting for the exploration of the higher one.
- Results also show comparable exploration for the Top-Down Guidance and the Goal Babbling control conditions, whereas the competent exploration measure shows that the architectures learning an intermediate hand's movement representation allow the agent to put again the object on much more diverse locations. An intuition for this result is that when exploring around motor parameters that lead to a movement of the block, the control architecture modifies directly joints' trajectories leading to less accurate collisions with the block than hierarchical architectures that try to modify hand's Cartesian trajectory. For that reason, **hierarchical architectures would produce less diverse but more reproducible interactions than control architectures**.
- In this setup, the intermediate hand's movement representation is reused by only one higher-level model, but **learning intermediate representations should be even more beneficial for exploration in complex learning hierarchies** where more than one models are reusing the learned representations.
- The **next step** is to integrate **social interaction** with a human peer that will give demonstrations on the task space and the hierarchy of models to explore.

References

[1] V. G. Santucci, G. Baldassarre, and M. Mirolli, "Which is the best intrinsic motivation signal for learning multiple skills?" *Frontiers in Neurorobotics*, 2013.

[2] A. Baranes and P.-Y. Oudeyer, "Active learning of inverse models with intrinsically motivated goal exploration in robots," *Robotics and Autonomous Systems*, vol. 61, no. 1, pp. 49–73, Jan. 2013.

[3] C. Moulin-Frier, P. Rouanet, P.-Y. Oudeyer, and others, "Explauto: an open-source Python library to study autonomous exploration in developmental robotics," in *ICDL-Epirob-International Conference on Development and Learning, Epirob*, 2014.

[4] N. Hansen, "The CMA evolution strategy: a comparing review," in *Towards a new evolutionary computation*. Springer, 2006.

[5] M. Lapeyre, P. Rouanet, J. Grizou, S. Nguyen, F. Depraetre, A. Le Falher, and P.-Y. Oudeyer, "Poppy Project: Open-Source Fabrication of 3D Printed Humanoid Robot for Science, Education and Art," in *Digital Intelligence 2014*, Nantes, France.

[6] A. J. Ijspeert, J. Nakanishi, H. Hoffmann, P. Pastor, and S. Schaal, "Dynamical movement primitives: learning attractor models for motor behaviors," *Neural computation*, vol. 25, no. 2, pp. 328–373, 2013.