Curiosity-driven Exploration of Skill Hierarchies

Sébastien Forestier
INRIA Bordeaux Sud-Ouest
Bordeaux, France
Email: sebastien.forestier@inria.fr

Pierre-Yves Oudeyer
INRIA Bordeaux Sud-Ouest
Bordeaux, France
Email: pierre-yves.oudeyer@inria.fr

***Abstract*—The abstract goes here.**

I. Introduction

The study of the control of manipulation actions in humans has revealed a modular representation of actions either in the cerebral cortex and in the spinal cord with compositionality: an infinite number of movements can be expressed through combination of simple primitives, and generalization: certain neurons (higher in the hierarchy) can represent actions independently of the effectors used [1].

The same idea holds for language expressiveness which is based on syntactic hierarchical combinations on a vocabulary, that open infinite semantic possibilities. Greenfield has also argued that this parallel between manipulation and language compositionality can be found in the human ontegenic development with combinatorial steps for manipulation and syntax acquired approximately at the same period and in the same order [2]. Also, the author explains that the development of the neural substrates for language and tool use could be an ontogenic homology as first of all the same neural computations for hierarchical combinations and their semantics should take place for both modalities, and furthermore experiments with Broca's and Wernicke's aphasics show that hierarchical organization for language and manipulation is linked. Broca's aphasics, who have less syntactic organization of speech were shown to also have problems of representation of the hierarchical organization of constructions with blocks, whereas Wernicke's aphasics, whose syntax is normal but speech semantics is impaired, succeed in representing such objects hierarchies.

Functional MRI experiments by Higuchi et al. have shown that the human's neural substrates for tool use and language is indeed shared in the dorsal BA44 Broca's area [3], which gives evidence for the similar neural computations used. They furthermore argue that these results supports the hypothesis that tool use have appeared first in primate evolution in F5 area, and then the language has developed in humans reusing part of tool use and manipulation neural substrates in human's Broca area, homolog of primate's F5.

Like a developing child, a developmental robot will have to incrementally explore skills that add up to the hierarchy of previously learned skills throughout its life, with a constraint being the cost and time of experimentation. We will seek to define curiosity-driven hierarchical learning architectures that could reuse the sensorimotor contingencies previously learned and to combine them to explore more efficiently new complex sensorimotor models.

A. Goal of the study

- Exploring in a structured hierarchy is more efficient than directly from M to S.
- Which task should I explore now ?
- How to choose between different means to explore a given space ?
- How can high-level tasks guide the exploration of lower-level ones ?
- How can the system cope with perturbations on some of the forward models ?

B. Related work

[4], [5].

Different computational models have the possibility to learn skill hierarchies. In finite environments represented by a factored Markov Decision Process [6], an intrinsic motivation towards actions maximizing Dynamic Bayesian Networks' structure has been shown to allow the learning of the environment's structure.

In continuous environments but with discrete actions, Metzen et al. [7] use the framework of options [8] to learn skill hierarchies. An intrinsic motivation rewards positively the novelty of the states encountered and negatively the prediction error of the learned skill model.

The model from Fabisch et al. [9] learns in a setting with a discrete task space (called contexts). It uses an intrinsic motivation for learning progress, and a Multi-Armed Bandit algorithm (D-UCB) to choose on which context the agent should train for. The Upper Confidence Bound algorithm chooses between contexts given their estimated learning progress and the uncertainty of these estimations by picking the context with the maximum upper confidence bound. In other words, it maximizes the expected reward plus something related to the uncertainty associated with it, selecting either contexts with certain high rewards or ones with uncertain poor reward. This algorithm embeds directly a solution the exploration-exploitation trade-off problem as it represents the exploitation of knowledge by the expected progress and the exploration of other solutions by the uncertainty bonus. This algorithm supposes a stationary learning progress on each context so the authors use an adaptation (D-UCB, [10]) to encompass non-stationary learning progress.

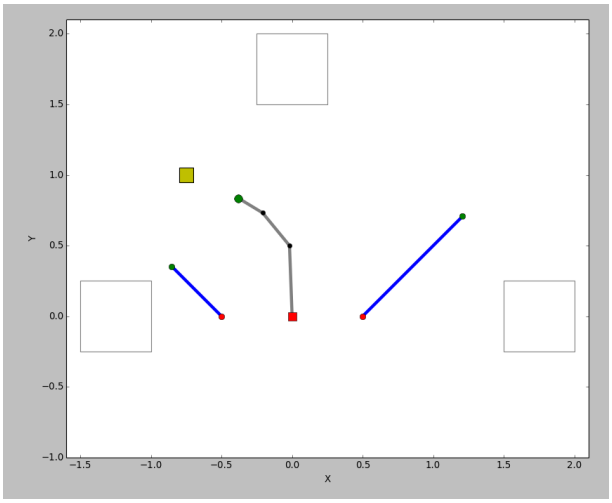


Fig. 1. Play Environment

In a fully continuous setting, Mugan et al. [11] have developed an algorithm that first learns a qualitative representation of environment states and actions in order to then learn the structure of Dynamic Bayesian Networks representing the temporal contingencies of those states and actions. In order to choose which action to practice, the authors use the IAC [12] where the agent is intrinsically motivated to choose actions that are estimated to yield high prediction error progress.

Here, we will rely more specifically on the SAGG-RIAC architecture [13]. This architecture learns a single mapping between continuous motor and sensori (or task) spaces with a competence-based intrinsic motivation. In our hierarchy of sensorimotor models, each model will be explored using the SAGG-RIAC procedure, but it could be replaced by another one without changing the mechanisms to learn the hierarchy that will be assessed.

II. Environment

See Fig. 1 for the initial state of the environment.

III. Random Goal Babbling

IV. Hierarchically structured exploration

A. Experiment 1: Methods

- Idea: to compare our simplest algorithm (explore the module with higher progress) in a realistic setting (Hierarchy (a) of Fig. **??**) to the control condition where a sensorimotor model is learned directly from the whole motor space M to the whole sensori space S.
- Conditions: Hierarchy (a) vs $M \rightarrow S$, Motor Babbling vs SAGG-Random
- Features: MAB on all modules, NN, No TDD.
- Measures: exploration of intermediate spaces (hands, tools), exploration of top spaces (objects). Competence to reach random goals in reachable parts of intermediate and top spaces. Statistics on multiple runs to see regularity/diversity in developmental trajectories.

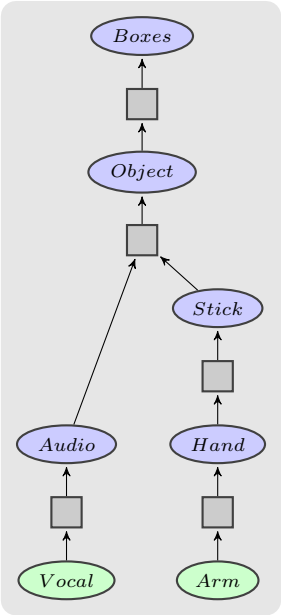


Fig. 2. H1

B. Experiment 1: Results

See Fig. 3, 4 and 5.

C. Experiment 1: Discussion

V. Choice of module to explore

A. Experiment 2: Methods

- Idea: to compare the different possibilities to choose the module to explore in the hierarchy: maximizing the progress, maximizing with a bias towards lower-level modules, or use ZPDES, with the same hierarchy (a) of Fig. **??**.
- Conditions: Random module, MAB on all modules, MAB with bias, ZPDES.
- Features: Hierarchy (a), SAGG-Random, NN, No TDD.
- Measures: exploration of intermediate spaces (hands, tools), exploration of top spaces (objects). Competence to reach random goals in reachable parts of intermediate and top spaces. Statistics on multiple runs to see regularity/diversity in developmental trajectories.

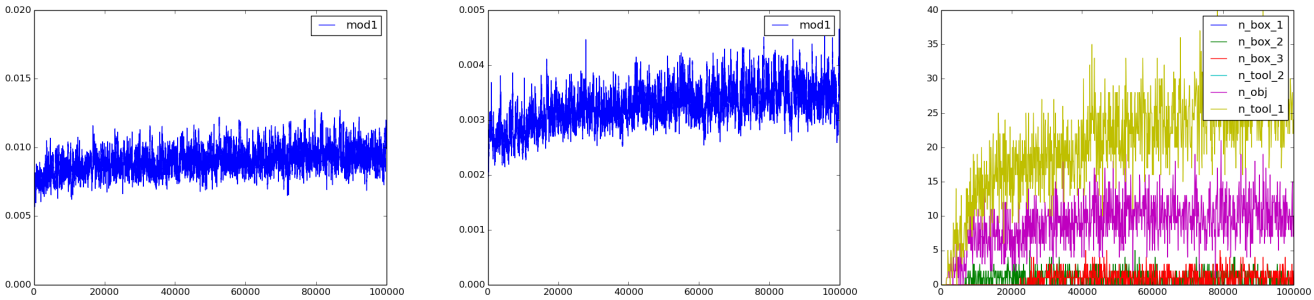
B. Experiment 2: Results

C. Experiment 2: Discussion

VI. Choice of tool to use

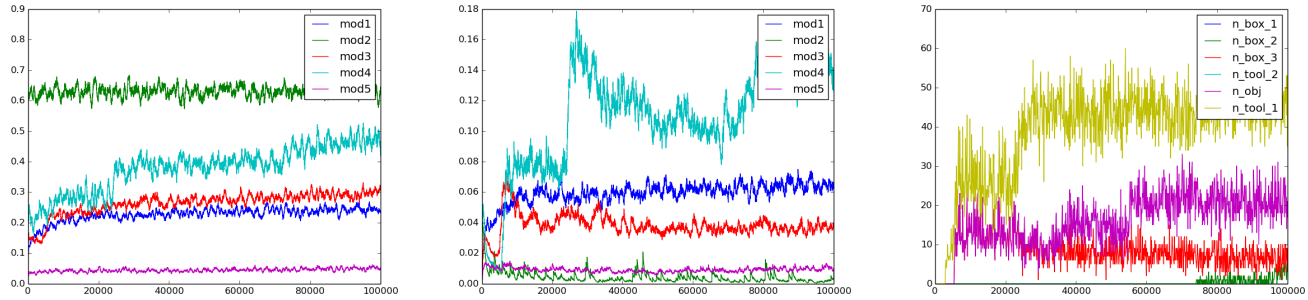
A. Experiment 3: Methods

- Idea: to explain how we can choose between different means (e.g. different tools) using the one with the maximal competence, or maximal progress. We can use the hierarchy (b) of Fig. **??**, in order to have 2 different tools to move the object.
- Conditions: maximize competence vs progress, noise on each tool, reachability (e.g. size) of each tool.



(a) Competences (b) Interests (c) Events

Fig. 3. Competences, interests and events using H0 with Random Goal Babbling



(a) Competences (b) Interests (c) Events

Fig. 4. Competences, interests and events using H1 with Random Goal Babbling

- Features: Hierarchy (b), MAB on all modules, SAGG-Random, NN, No TDD.
- Measures: exploration of intermediate spaces (hand, tools), exploration of top spaces (object). Competence to reach random goals in reachable parts of intermediate and top spaces. Statistics on multiple runs to see regularity/diversity in developmental trajectories.

B. Experiment 3: Results

C. Experiment 3: Discussion

VII. Top-Down Guidance

A. Experiment 4: Methods

- Idea: to compare different possibilities of Top-Down Guidance, with hierarchy (c) of Fig. **??** in order to have TD guidance at different levels: arm with one higher model, hand with 2 higher models, tool1 with one higher model, or tool2 with 2 higher models.
- Conditions: No TDD, Just add noise to motor command of each module (pb: interferes with competence estimation) vs explore n points and returns the best (warning: exponential) vs only one layer below the babbling module explores n points.
- Features: Hierarchy (c), MAB on all modules with bias (or no bias?), SAGG-Random, NN.
- Measures: exploration of intermediate spaces (hands, tools), exploration of top spaces (objects). Competence to reach random goals in reachable parts of intermediate and top spaces. Statistics on multiple runs to see regularity/diversity in developmental trajectories.

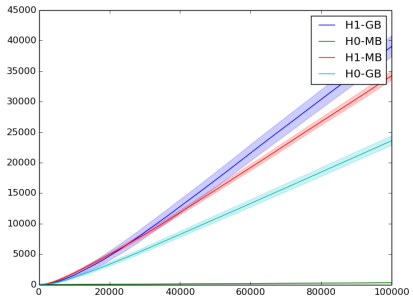
B. Experiment 4: Results

C. Experiment 4: Discussion

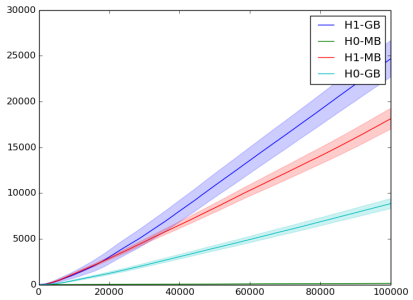
VIII. Robustness to perturbations

A. Experiment 5: Methods

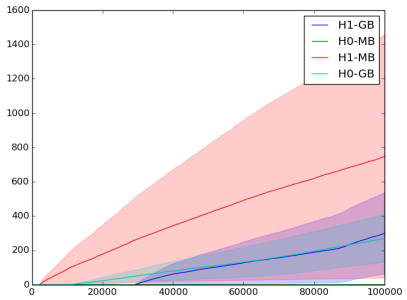
- Idea: to apply perturbations to one of the possible forward models, either blocking, shifting or randomizing one dimension. We can use the hierarchy (d) of Fig. **??** in order to see an adaptation at 2 levels: the use of one hand to use one tool, the use of both tools to move the object even if only one is perturbated.
- Conditions: No perturbations, which model is perturbated (arm, tool), type of perturbation (blocking, shifting, random).
- Features: Hierarchy (d), MAB on all modules, SAGG-Random, NN, best TDD.
- Measures: exploration of intermediate spaces (hands, tools) and top spaces (objects) before and after perturbations. Competence to reach random goals in reachable parts of intermediate and top spaces before and after



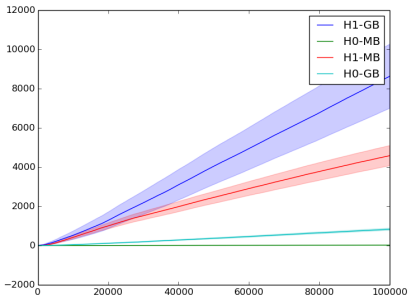
(a) Tool



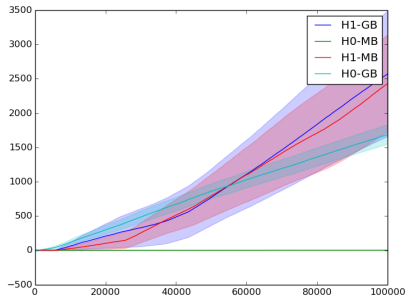
(b) Object



(c) Box 1



(d) Box 2



(e) Box 3

Fig. 5. Number of touch of tool, object and boxes for each 100 iterations' bin.

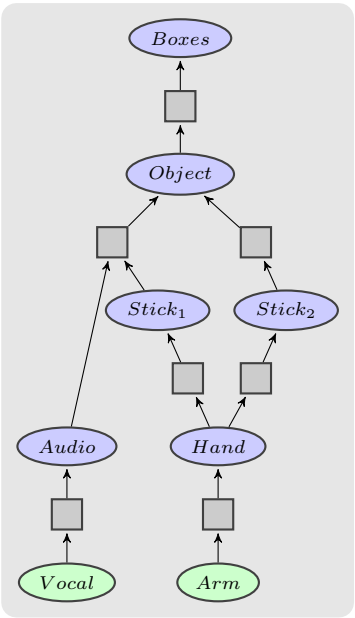


Fig. 6. H2

perturbations. Statistics on multiple runs to see regularity/diversity in developmental trajectories.

B. Experiment 5: Results

C. Experiment 5: Discussion

IX. General Discussion

References

[1] A. Cangelosi, G. Metta, G. Sagerer, S. Nolfi, C. Nehaniv, K. Fischer, J. Tani, T. Belpaeme, G. Sandini, F. Nori *et al.*, "Integration of action and language knowledge: A roadmap for developmental robotics," *Autonomous Mental Development, IEEE Transactions on*, vol. 2, no. 3, pp. 167–195, 2010.

[2] P. M. Greenfield, "Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior," *Behavioral and Brain Sciences*, vol. 14, pp. 531–551, 12 1991.

[3] S. Higuchi, T. Chaminade, H. Imamizu, and M. Kawato, "Shared neural correlates for language and tool use in broca's area," *Neuroreport*, vol. 20, no. 15, pp. 1376–1381, 2009.

[4] E. Ugur and J. Piater, "Emergent structuring of interdependent affordance learning tasks," in *Development and Learning and Epigenetic Robotics (ICDL-Epirob), 2014 Joint IEEE International Conferences on*. IEEE, 2014, pp. 489–494.

[5] E. Ugur, Y. Nagai, E. Sahin, and E. Oztop, "Staged development of robot skills: Behavior formation, affordance learning and imitation with motionese," *Autonomous Mental Development, IEEE Transactions on*, vol. 7, no. 2, pp. 119–139, June 2015.

[6] C. M. Vigorito and A. G. Barto, "Intrinsically motivated hierarchical skill learning in structured environments," *Autonomous Mental Development, IEEE Transactions on*, vol. 2, no. 2, pp. 132–143, 2010.

[7] J. H. Metzen and F. Kirchner, "Incremental learning of skill collections based on intrinsic motivation," *Frontiers in Neurorobotics*, vol. 7, 2013.

[8] R. S. Sutton, D. Precup, and S. Singh, "Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning," *Artificial intelligence*, vol. 112, no. 1, pp. 181–211, 1999.

[9] A. Fabisch and J. H. Metzen, "Active contextual policy search," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 3371–3399, 2014.

[10] L. Kocsis and C. Szepesvári, "Discounted ucb." 2nd PASCAL Challenges Workshop, 2006.
[11] J. Mugan and B. Kuipers, "Autonomously learning an action hierarchy using a learned qualitative state representation," 2009.
[12] P.-Y. Oudeyer, F. Kaplan, and V. V. Hafner, "Intrinsic Motivation Systems for Autonomous Mental Development," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 2, pp. 265–286, Apr. 2007.
[13] 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.