developmental active learning with intrinsic motivation

Pierre-Yves Oudeyer, Adrien Baranes

Developmental active learning with intrinsic motivation

Pierre-Yves Oudeyer and Adrien Baranès
INRIA Bordeaux - Sud-Ouest
pierre-yves.oudeyer@inria.fr, adrien.baranes@inria.fr

Developmental learning in unprepared sensorimotor spaces

In developmental robotics (Weng et al., 2001; Lungarella et al., 2003), one aims to investigate the mechanisms that may allow a robot to continuously discover and learn new skills in unknown environments and in a life-long time scale. Of particular importance is the fact that the set of these skills and their function(s) are at least partially unknown to the engineer who conceive the robot initially, and are also task-independent. Indeed, a desirable feature is that robots should be capable of exploring and developing various kinds of skills that they may re-use later on for tasks that they did not foresee. This is what happens in human children, and this is also why developmental robotics shall import concepts and mechanisms from human developmental psychology.

Like children, the "freedom" that is given to developmental robots to learn an open set of skills also poses a very important problem: as soon as the set of motors and sensors is rich enough, the set of potential skills become extremely large and complicated. This means that on the one hand, it is impossible to try to learn all skills that may potentially be learnt because there is not enough time, and also that there are many skills or goals that the child/robot could imagine but never be actually learnable, because they are either too difficult or just not possible (for example, trying to learn to control the weather by producing gestures is hopeless). This kind of problem is not at all typical of the existing work in machine learning, where usually the "space" and the associated "skills" to be learnt and explored are well-prepared by a human engineer. For example, when learning hand-eye coordination in robots, the right input and output spaces (e.g. arm joint parameters and visual position of the hand) are typically provided as well as the fact that hand-eye coordination is an interesting skill to learn. But a developmental robot is not supposed to be provided with the right subspaces of its rich sensorimotor space and with their association with appropriate skills: it would for example have to discover that arm joint parameters and visual position of the hand are related in the context of a certain skill (which we call hand-eye coordination but which it has to conceptualize by itself) and in the middle of a complex flow of values in a richer set of sensations and actions.

Intrinsic motivation

Thus, developmental robots have a sharp need for mechanisms that may drive and self-organize the exploration of new skills, as well as identify and organize useful sub-spaces in its complex sensorimotor experiences. In psychology terms, this amount to trying to answer the question "What is interesting for a curious brain?". Among the various trends of research which have approached this question, of particular interest is work on intrinsic motivation. Intrinsic motivations are mechanisms that guide curiosity-driven exploration, that were initially studied in psychology (White, 1959; Berlyne, 1965; Deci et Ryan, 1985) and are now also being approached in neuroscience (Schultz, 2002; Dayan and Balleine, 2002; Redgrave and Gurney, 2006). They have been proposed to be crucial for self-organizing developmental trajectories (Oudeyer et al. , 2007) as well as for guiding the learning of general and reusable skills (Barto et al., 2005). Experiments have been conducted in real-world robotic setups, such as in (Oudeyer et al., 2007) where an intrinsic motivation system was shown to allow for the progressive discovery of skills of increasing complexity, such as reaching, biting and simple vocal imitation with and AIBO robot. In these experiments, the focus was on the study of how developmental stages could self-organize into a developmental trajectory without a direct pre-specification of these stages and their number.

Developmental active learning

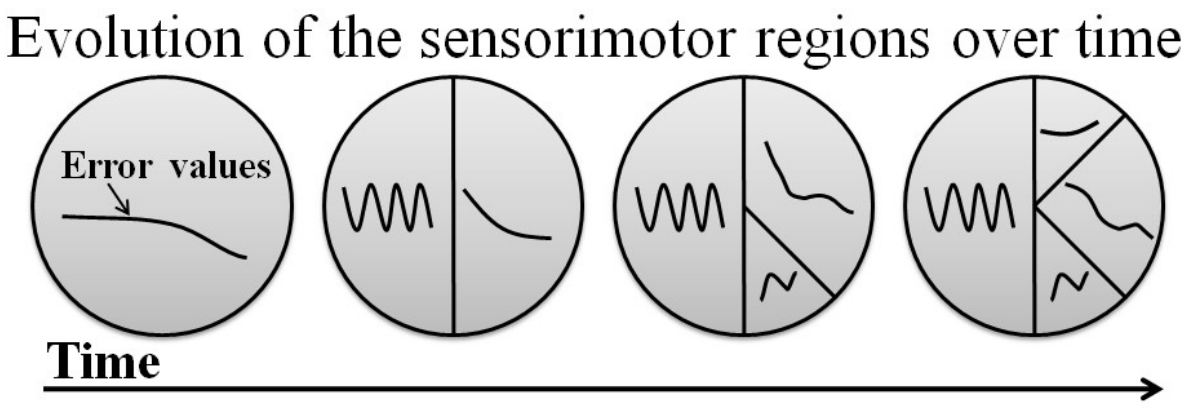
Here, we argue that they can also be considered as "active learning" algorithms, and show that some of them also allow for very efficient learning in the unprepared spaces with the typical properties of those encountered by developmental robots, outperforming standard active learning heuristics.

One essential activity of epigenetic robots is to learn forward models of the world, which boils down to learning to predict the consequences of its actions in given contexts. This learning happens as the robot collects learning examples from its experiences. If the process of example collection is disconnected from the learning mechanism, this is called passive learning. In contrast, researchers in machine learning have proposed algorithms allowing the machine to choose and make experiments that maximize the expected information gain of the associated learning example (Cohn et al., 1996), which is called "active learning". This has been shown to dramatically decrease the number of required learning examples in order to reach a given performance in data mining experiments (Hasenjager and Ritter, 2002), which is essential for a robot since physical action takes time. The typical active learning heuristics consist in focusing the exploration in zones where unpredictability or uncertainty of the current internal model are maximal, which involves the online learning of a meta-model that evaluates this unpredictability or uncertainty. In the following, we will implement a version of this heuristics which we will denote "MAX", and consists in an ε-greedy strategy in which the experiment for which the meta-model predicts the highest prediction error of the model is chosen with a probability p, and a uniformly random experiment is chosen with a probability $1 - p$.

Unfortunately, it is not difficult to see that it will fail completely in unprepared robot sensorimotor spaces. Indeed, the spaces that epigenetic robots have to explore are typically composed of unlearnable subspaces, such as for example the relation between its joints values and the motion of unrelated objects that might be visually perceived. Classic active learning heuristics will push the robot to concentrate on these unlearnable zones, which is obviously undesirable.

Based on psychological theories proposing that exploration is focused on zones of optimal intermediate difficulty or novelty (Berlyne, 1960), intrinsic motivation mechanisms have been proposed, pushing robots to focus on zones of maximal learning progress (see Oudeyer et al., 2007 for a review). As exploration is here closely coupled with learning, this can be considered as active learning. Through a number of systematic experiments on artificially generated mappings that include unlearnable and inhomogeneous zones, we argue that this kind of intrinsically motivated exploration actually permits organized and very efficient learning, vastly outperforming standard active learning methods.

In the presented system, "interesting" experiments are defined as those where the predictions improve maximally fast, hence the term "learning progress". In order to compute and predict learning progress (this is in fact a meta-prediction), (Oudeyer et al., 2007) introduced the concept of "regions" which are sub-spaces of the sensorimotor space, recursively and progressively defined, to each of which is attached a global interest value, which is the inverse of the global mean prediction error derivative in the past in the region. This system of regions and their associated learning progress values constitute the meta-model of the active learning system, and it should be noted that these regions do not necessarily reflect the internal organization of the model (though it can, as in (Oudeyer et al., 2007)). The algorithm starts from a single large region (the whole space), which it progressively subdivides in such a way that the dissimilarity of each sub-region in terms of learning progress is maximal. Here is an illustration about this recursive parsing system, over time:



Based on this partitioning and associated evaluation of interest, the following ***ε-greedy exploration policy*** is used when a new sensorimotor experiment has to be chosen:

- **(Meta-exploitation)** With a probability p (typically equal to 0.7), a sensorimotor experiment is uniformly randomly chosen in the region which has the *highest associated learning progress*;
- **(Meta-exploration)** With a probability $1 - p$ (typically equal to 0.3), then:
    - With a probability 0.5, choose uniformly randomly an experiment;
    - With a probability 0.5, choose an experiment using the MAX heuristics;

The meta-exploration part is indeed necessary to allow the system to discover niches of learning progress: any region needs to be explored a little bit first in order to let the system know how much it is interesting or not.

Experiments

We now compare the performance of this system, denoted IAC for Intelligent Adaptive Curiosity, when viewed as an active learning algorithm and compared to the MAX heuristics and to the naïve RANDOM heuristics (which consists in choosing uniformly randomly experiments). An experiment was conducted in an abstract space characterized by properties typical of the unprepared spaces that might be encountered by a developmental robot: simple zones, more complex zones, and unlearnable zones. This space is a $R^1 \rightarrow R^1$ sensorimotor space which incorporates areas of different difficulty:

- The interval [0.25, 0.5] ∪ [0.9, 1] contains an unlearnable situation (pure noise)
- [0.5, 0.8] is an increasing difficulty area
- [0.15, 0.17] an intermediate complexity part
- The rest is easy to learn for the algorithm

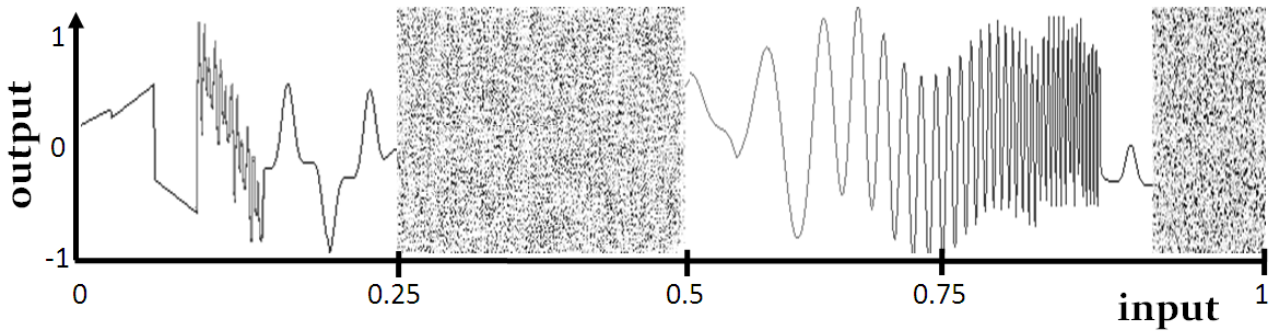


Figure 1 The abstract sensorimotor space ($R^1 \rightarrow R^1$) to be learnt. It was built to capture some of the important peculiar features of the typical sensorimotor spaces that developmental robots will encounter: zones of various levels of difficulty and even unlearnable zones. It was here deliberately chosen to be low-dimensional (which is not a typical feature of the spaces encountered by developmental robots) so that the behavior of the IAC heuristics could be visualized and studied thoroughly in relation to its capacity to focus on zones of increasing complexity and at the same time avoid unlearnable zones.

The next figure (Figure 2) shows the evolution of exploration focus of IAC over time, when the system is driven by intrinsic motivation. We see that it avoids unlearnable zones, yet focusing on the difficult parts of the learnable zones:

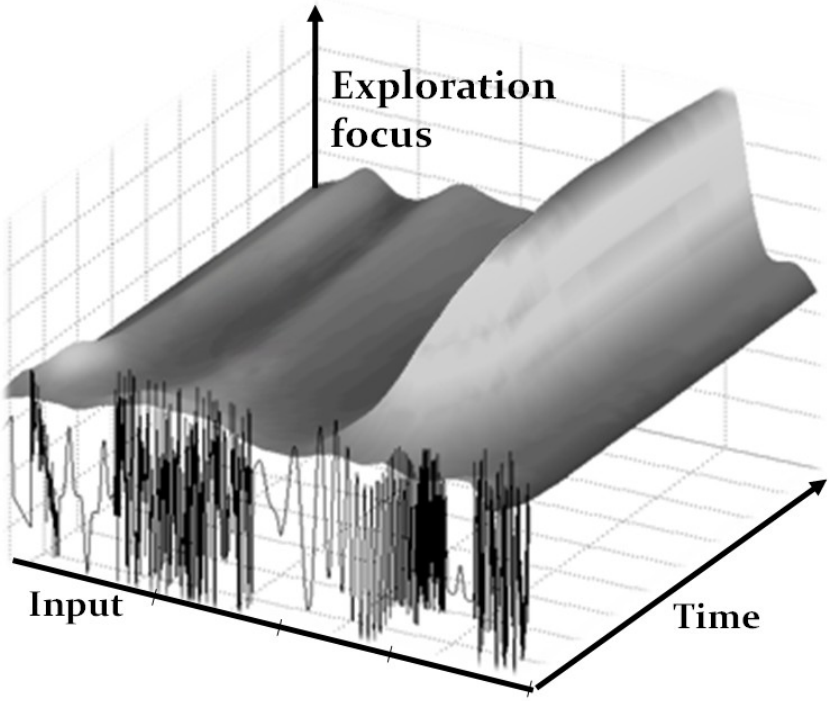


Figure 2 Exploration focus over time

We can notice, watching the previous graph in the start-up phase, that the algorithm is interested in the noisy part [0.25, 0.5], but only during a brief period; then, it changes of area and decide to focus on the first complex one, [0.15, 0.17]. Finally, it focuses on the zone within [0.7, 0.8], beginning in its most simple sub-part, and progressively shifting to its more difficult part.

Thereby, we see that the system is not trapped in unlearnable zones, as opposed to traditional active learning methods, but still focuses on zones where effort is most needed, as for random exploration.

In figure 3, we compare the evolution of performance in generalization inside learnable zones using the error rate, among IAC, MAX and RANDOM. To remain fair, the MAX heuristics was implemented with the same 0.7 meta-exploitation/0.3 meta-exploration global scheme than IAC (the difference is thus in the meta-exploitation part).

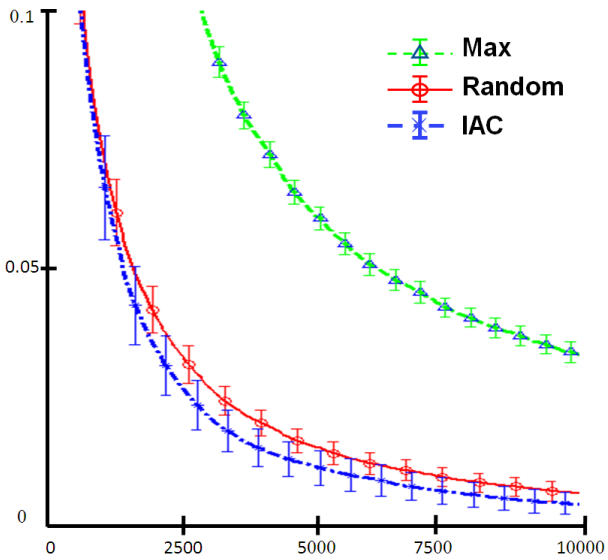


Figure 3 Evolution of performances in generalization

We obtain better performances in learning using the IAC algorithm. We observe that over time, error rate is always inferior to others. This shows that building an active learning system based on intrinsic motivation and developmental concepts coming from psychology, one can obtain better learning performances in unprepared sensorimotor spaces than traditional active learning techniques.

References

Barto, A., Singh S., and Chentanez N. (2004) Intrinsically motivated learning of hierarchical collections of skills, in Proc. 3rd Int. Conf. Development Learn., San Diego, CA, 2004, pp. 112–119.

Berlyne, D. (1960). Conflict, Arousal, and Curiosity. New York: McGraw-Hill.

Cohn D., Ghahramani Z., and Jordan M. (1996) Active learning with statistical models, J. Artif. Intell. Res., vol. 4, pp. 129–145, 1996.

Deci, E. and Ryan, R. (1985). Intrinsic Motivation and Self-Determination in Human Behavior. Plenum Press.

Dayan, P. and Balleine, B. (2002) Reward, Motivation and Reinforcement Learning, Neuron, Vol. 36, pp. 285-298.

Hasenjager M.and Ritter H. (2002) Active Learning in Neural Networks. Berlin, Germany: Physica-Verlag GmbH, Physica-Verlag Studies In Fuzziness and Soft Computing Series, pp. 137–169.

Lungarella, M., Metta, G., Pfeifer, R., and Sandini, G. (2003) Developmental robotics: A survey, Connection Sci., vol. 15, no. 4, pp. 151–190, 2003.

Oudeyer P-Y & Kaplan , F. & Hafner, V. (2007) Intrinsic Motivation Systems for Autonomous Mental Development, IEEE Transactions on Evolutionary Computation, 11(2), pp. 265--286.

Redgrave, P. and Gurney, K. (2006) The Short-Latency Dopamine Signal: a Role in Discovering Novel Actions?, Nature Reviews Neuroscience, Vol. 7, no. 12, pp. 967-975.

Schultz, W. (2002) Getting Formal with Dopamine and Reward, Neuron, Vol. 36, pp. 241-263.

Weng, J., McClelland J., Pentland, A., Sporns, O., Stockman, I., Sur M., and Thelen, E. (2001) Autonomous mental development by robots and animals, Science, vol. 291, pp. 599–600.

White, R. (1959). Motivation reconsidered: The concept of competence. Psychological review, 66:297–333.