Intrinsically motivated exploration as efficient active learning in unknown and unprepared spaces

Pierre-Yves Oudeyer, Adrien Baranes

Intrinsically motivated exploration as efficient active learning in unknown and unprepared spaces

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

Intrinsic motivations are mechanisms that guide curiosity-driven exploration (Berlyne, 1965). 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). Here, we argue that they can be considered as "active learning" algorithms, and show that some of them also allow for very efficient learning in unprepared sensorimotor spaces, outperforming existing active learning algorithms.

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 is maximal. In the following, we will implement a version of this heuristics which we will denote "MAX".

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 spaces, 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. 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:

Evolution of the sensorimotor regions over time
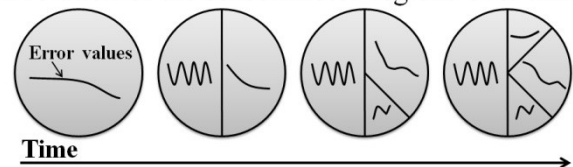
Based on this partitioning and associated evaluation of interest, the following *exploration policy* is used when a new sensorimotor experiment has to be chosen:

- **(Meta-exploitation)** With a probability 0.7, a sensorimotor experiment is uniformly randomly chosen in the region which has the highest associated learning progress;
- **(Meta-exploration)** With a probability 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.

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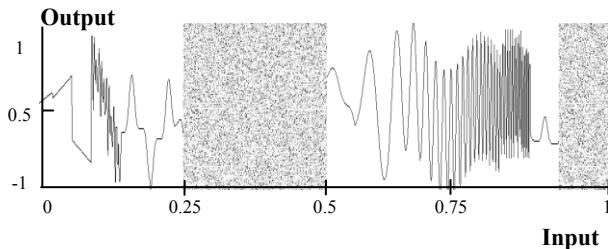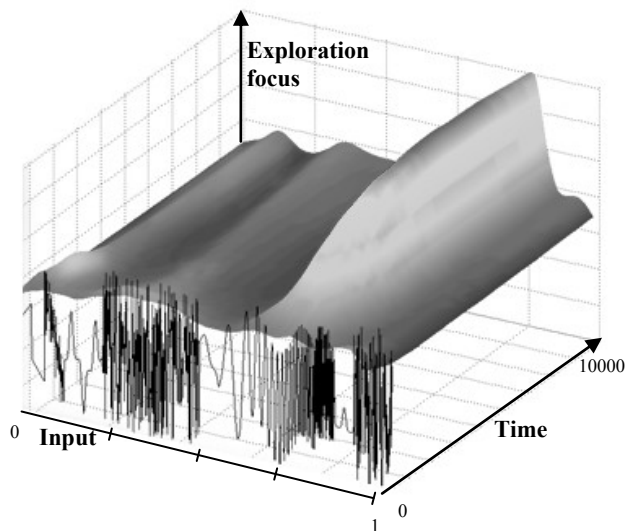- The rest is easy to learn for the algorithm

Figure 1: The abstract sensorimotor space, an input/output mapping to be learnt

The next figure (Graph a) 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:


Graph (a): 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 graph (b), 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).
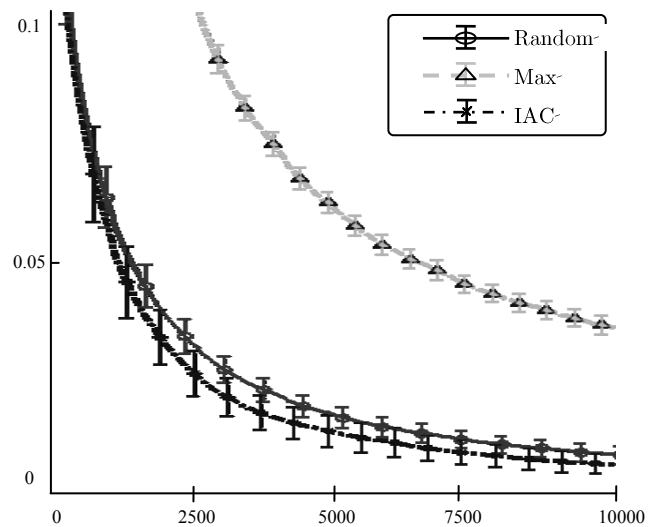
Graph (b): Evolution of performances in generalization

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.

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.

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.