Socially Guided Intrinsically Motivated Learner

Sao Mai Nguyen, Pierre-Yves Oudeyer

Socially Guided Intrinsically Motivated Learner

Sao Mai Nguyen[1] and Pierre-Yves Oudeyer[1]

***Abstract*— This paper studies the coupling of two learning strategies: internally guided learning and social interaction. We present Socially Guided Intrinsic Motivation by Demonstration (SGIM-D) and its interactive learner version Socially Guided Intrinsic Motivation with Interactive learning at the Meta level (SGIM-IM), which are algorithms for learning inverse models in high dimensional continuous sensorimotor spaces. After describing the general framework of our algorithms, we illustrate with a fishing experiment.**

I. INTRODUCTION

In initial work to address multi-task learning, we proposed the Socially Guided Intrinsic Motivation by Demonstration (SGIM-D) algorithm [1] which merges socially guided exploration as defined in [2], [3] and intrinsic motivation [4], [5] based on SAGG-RIAC algorithm [6], to reach goals in a continuous task space, in the case of a complex, high-dimensional and continuous environment. While the SGIM-D learner passively uses demonstrations given by a teacher at regular frequency, the **SGIM-IM** (Socially Guided Intrinsic Motivation with Interactive learning at the Meta level) algorithm [7] optimises the timing of the interactions with the teacher and actively chooses between autonomous and social learning strategies.

A. Formalisation

Let us consider an agent who can complete tasks τ parameterised by $\theta \in T$, by carrying out policies $\pi_u : A \to [0, 1]$, parameterised by $u \in \Pi$.

The performance of a policy π_u at completing a task τ_θ is measured by:

$$\begin{aligned} J : & T \times \Pi && \to [0, +\infty[\\ & (\theta, u) && \mapsto J(\theta, u) \end{aligned} \quad (1)$$

We define a $skill$ as the function that maps to a task τ the best policy to complete it:

$$\begin{aligned} S : & T && \to \Pi \\ & \theta && \mapsto argmax_u J(\theta, u) \end{aligned} \quad (2)$$

We assume that T can be partitioned into subspaces where the tasks are related, and in these subspaces our parametrisation allows a smooth variation of J with respect to θ most of the time, i.e. that S is a piecewise continuous function. The aim of the agent is to find the right policy to complete every task τ_θ to maximise $I = \int_\theta J(\theta, S(\theta)) d\theta$ by self-exploring the policy and task spaces and by asking for help to a teacher, who performs a trajectory ζ_d and completes a task τ_{θ_d}.

[1]Flowers Team, INRIA and ENSTA ParisTech, France. nguyensmai at gmail.com, pierre-yves.oudeyer at inria.fr

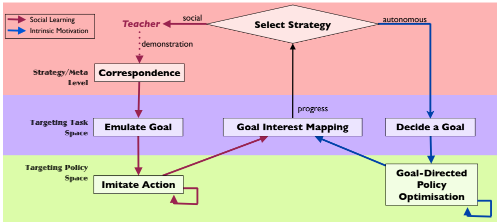


Fig. 2: Time flow chart of SGIM-D and SGIM-IM, which combines Intrinsic Motivation and Social Learning into 3 layers.

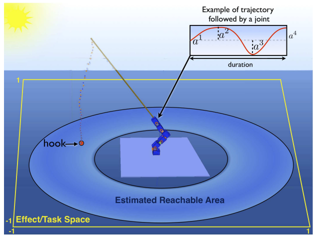


Fig. 3: Fishing experimental setup.

B. Algorithm Outline

SGIM-D and SGIM-IM learn by episodes (fig. 1) during which they actively choose a task $\tau_{\theta_g} \in T$ to reach with a intrinsically motivated exploration or imitation strategy (fig. 2). The interactive learner SGIM-IM also chooses a learning strategy, based on the progress made by each of them.

In an episode under the intrinsic motivation strategy, the learner explores autonomously following the SAGG-RIAC algorithm [6]. It actively self-generates a goal τ_g where its competence improvement is maximal, then explores which policy π_u can achieve τ_{θ_g} best. The SGIM-D and SGIM-IM learners explore preferentially goal tasks easy to reach and where it makes progress the fastest. It tries different policies to approach the self-determined task τ_{θ_g}, re-using and optimising the policy built through its past autonomous and socially guided explorations. The episode ends after a fixed duration.

In an episode under the imitation strategy, the learner observes from the selected teacher a demonstration $[c_d, \zeta_d, \tau_d]$, memorises this effect τ_d as a possible goal, and mimics the teacher by performing policies π_u to reproduce ζ_d, for a fixed duration.

The architectures of the SGIM algorithms are separated into 3 levels: task space exploration, policy space exploration and strategy selection. Pseudo-codes are detailed in [1], [7].

II. FISHING EXPERIMENT

A. Experimental Setup

A 6-dof arm learns how to place the hook at the tip the fishing line at any point on the surface of the water (fig 3). It

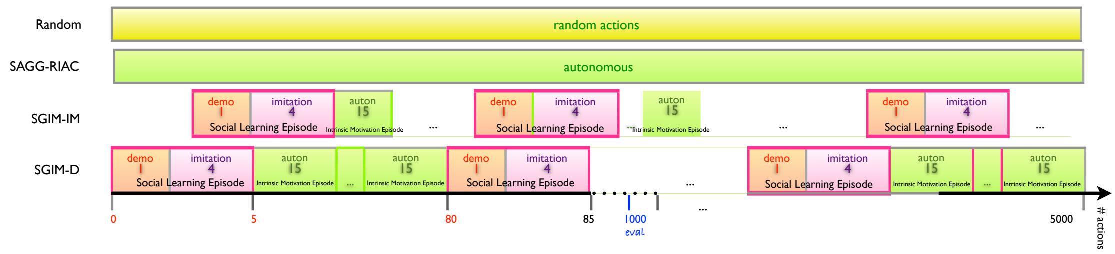


Fig. 1: Comparison of several learning algorithms. Each box represents the chronology of the adopted strategies (the figures correspond to the number of actions experimented in the episode). The figures here are given for the Fishing experiment).

learns high-dimensional models between 25 and 2-D spaces, for highly redundant problems, as detailed in [1].

B. Results

The SGIM algorithms manage to take advantage of the properties of the demonstrations [8] to bootstrap its autonomous exploration in order to:

- complete most tasks with higher precision (fig. 4) . SGIM-IM and SGIM-D make smaller error than random or SAGG-RIAC.
- explore more tasks (fig. 5). SGIM-D and SGIM-IM complete more tasks.

The interactive learner SGIM-IM could also balance leaning by imitation and autonomous learning, by taking into account its progress with each of the strategies, and the cost of an interaction, so as to minimise the teacher's effort and maximise the impact of each demonstration (fig. 6).

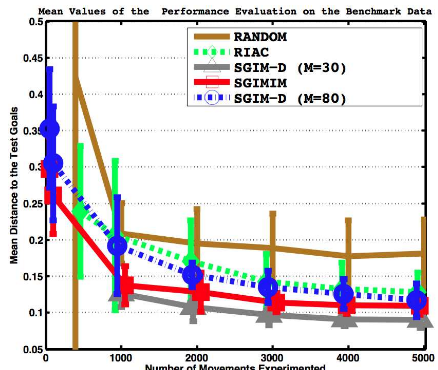


Fig. 4: Evaluation of the performance of the robot with respect to the number of actions performed, under several learning algorithms. We plotted the mean distance with its standard deviation errorbar.

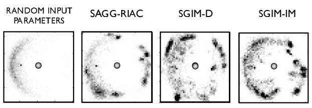
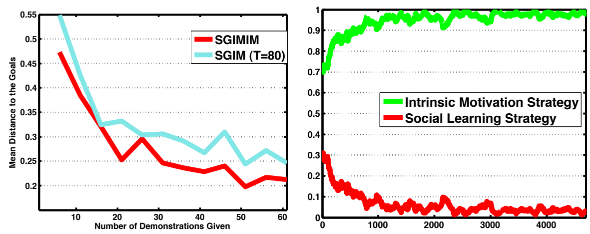


Fig. 5: Histogram of the tasks explored by the fishing rod inside the 2D effects space. .

III. Discussion and Conclusion

The Socially Guided Intrinsic Motivation algorithms have hierarchical structures which include two levels of active learning. Based on its exploration in the action space, they actively choose in the task space which goals could be interesting to target. Based on the progress of each strategy,



(a) Comparison of the performance of the robot with respect to the number of demonstrations given, of SGIM-IM and SGIM-D

(b) Strategy chosen though time: percentage of times each strategy is chosen with respect to the number of actions performed

Fig. 6: Strategy Selection of SGIM-IM.

SGIM-IM selects on a meta level between autonomous learning or social learning strategies. The learner can actively interact with the teacher. This structure can be extended to take into account more complex social interaction scenarios, such as an interaction with several teachers, where the learner can choose who it should imitate. Future work will study possibilities for the robot to request for specific demonstrations (show me a specific kind of movements or show me how to complete a kind of goals).

Acknowledgment

This research was partially funded by ERC Grant EXPLORERS 240007 and ANR MACSi.

References

[1] S. M. Nguyen, A. Baranes, and P.-Y. Oudeyer, "Bootstrapping intrinsically motivated learning with human demonstrations," in *Proceedings of the IEEE International Conference on Development and Learning*, Frankfurt, Germany, 2011.

[2] A. Whiten, "Primate culture and social learning," *Cognitive Science*, vol. 24, no. 3, pp. 477–508, 2000.

[3] A. Billard, S. Calinon, R. Dillmann, and S. Schaal, *Handbook of Robotics*. MIT Press, 2007, no. 59, ch. Robot Programming by Demonstration.

[4] E. Deci and R. M. Ryan, *Intrinsic Motivation and self-determination in human behavior*. New York: Plenum Press, 1985.

[5] P.-Y. Oudeyer, F. Kaplan, and V. Hafner, "Intrinsic motivation systems for autonomous mental development," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 2, pp. 265–286, 2007.

[6] A. Baranes and P.-Y. Oudeyer, "Intrinsically motivated goal exploration for active motor learning in robots: A case study," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, oct. 2010, pp. 1766 –1773.

[7] S. M. Nguyen and P.-Y. Oudeyer, "Interactive learning gives the tempo to an intrinsically motivated robot learner," in *IEEE-RAS International Conference on Humanoid Robots*, 2012.

[8] ——, "Properties for efficient demonstrations to a socially guided intrinsically motivated learner," in *21st IEEE International Symposium on Robot and Human Interactive Communication*, 2012.