Socially Guided Intrinsically Motivated Learner
Sao Mai Nguyen, Pierre-Yves Oudeyer

To cite this version:
Sao Mai Nguyen, Pierre-Yves Oudeyer. Socially Guided Intrinsically Motivated Learner. IEEE International Conference on Development and Learning, Nov 2012, San Diego, United States. 10.1109/DevLrn.2012.6400809 . hal-00936960

HAL Id: hal-00936960
https://hal.archives-ouvertes.fr/hal-00936960
Submitted on 27 Jan 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
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 \(\tau\) parameterised by \(\theta \in T\), by carrying out policies \(\pi_u : A \rightarrow [0, 1]\), parameterised by \(u \in \Pi\).

The performance of a policy \(\pi_u\) at completing a task \(\tau_0\) is measured by:

\[
J : T \times \Pi \rightarrow [0, +\infty[ \\
(\theta, u) \mapsto J(\theta, u)
\]

We define a skill as the function that maps to a task \(\tau\) the best policy to complete it:

\[
S : T \rightarrow \Pi \\
\theta \mapsto \text{argmax}_u J(\theta, u)
\]

We assume that \(T\) can be partitioned into subspaces where the tasks are related, and in these subspaces our parameterisation allows a smooth variation of \(J\) with respect to \(\theta\) 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 \(\tau_0\) 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 \(\zeta_d\) and completes a task \(\tau_{d,t}\).

---

\(^1\)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.

B. Algorithm Outline

SGIM-D and SGIM-IM learn by episodes (fig. 1) during which they actively choose a task \(\tau_{0,t} \in T\) to reach with an 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 \(\tau_\pi\) where its competence improvement is maximal, then explores which policy \(\pi_u\) can achieve \(\tau_\pi\) 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 \(\tau_{0,t}\), 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 \(\tau_d\) as a possible goal, and mimics the teacher by performing policies \(\pi_u\) to reproduce \(\zeta_d\) 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
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).

(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


