Data Collection of an Interactive Learner for Strategy, Outcome and Policy Exploration

Sao Mai Nguyen and Pierre-Yves Oudeyer
Flowers Team, INRIA Bordeaux, France

***Abstract*—SGIM (Socially Guided Intrinsic Motivation) is a hierarchical learning architecture for robot control. It allows the robot to learn to complete various outcomes with several strategies in stochastic high-dimensional environments. More precisely, this interactive learner can actively choose to learn with strategies where it explores the environment autonomously guided by intrinsic motivation, or with social guidance. This algorithmic architecture is hierarchically structured into three active learning layers: policy space exploration, outcome space exploration, and strategy selection. The SGIM active learner automatically discovers and exploits the structure of the environment structure as well as the properties of the teachers. It allows automatic selection of the best teachers for a given task, and automatic discovery of the manifolds of the outcome space to allow generalisation (skill transfer) and exploration of new outcomes.**

I. Strategic Learning for Life-Long Learning in Real-World Environments

For life-long learning of multiple skills by robots, because of the cost of data acquisition, the learning agent has to use an efficient data collection strategy. We study how an active learning agent can achieve varied outcomes in structured continuous outcome spaces and how he can learn for those various outcomes which strategy to adopt.

II. A Hierarchical Algorithmic Architecture

Details of the architecture can be found in [3].

A. Episodic Reinforcement Learning

SGIM has to choose actively and hierarchically at each learning episode: 1) which goals in an outcome space to aim; 2) which data collection strategy to use (self-exploration/imitation); 3) if he chooses imitation, he chooses who to imitate.

B. 3-Layered Architecture

Its hierarchical architecture bears three levels. The lower level explores the policy parameters space to build skills for determined goal outcomes. The upper level explores the outcome space to evaluate for which outcomes he makes the best progress. A meta-level actively chooses the outcome and data collection strategy that leads to the best competence progress.

III. Skill Transfer, Structure of the Environment and Teachers

A. Skill Transfer

We showed in [3] that SGIM can focus on the outcome where it learns the most, and in [2, 3] that it can choose the most appropriate associated data collection strategy to complete new outcomes. The active learner can explore efficiently a composite and continuous outcome space to be able to generalise for new outcomes of the outcome spaces.

B. Discovery and Exploitation of the Structure of the Environment

We showed in [4, 3] that SGIM can also distinguish the reachable space from unreachable space, and can structure its exploration according to the levels of difficulty of the different outcomes, even in stochastic, continuous and high-dimensional environments like the fishing experiment. In [1], we showed that the properties of human demonstrations helped the learner to structure both his policy and outcome space exploration, even in the case of small discrepancies between the teacher and the learner.

C. Discovery and Exploitation of the Properties of the Teachers

In [3], we showed that it takes efficient advantage of the available teachers (with different levels of skills) to improve its learning performance and coherently selects the best strategy with respect to the chosen outcome. It is robust even to large discrepancies between the teacher and the learner.

IV. Conclusion

The **SGIM** algorithmic architecture efficiently and actively combines autonomous self-exploration and interactive learning, to address the learning of multiple outcomes, with outcomes of different types, and with different data collection strategies. It learns when to ask for demonstration, what kind of demonstrations (action to mimic or outcome to emulate) and who to ask for demonstrations, among the available teachers.

References

[1] Sao Mai Nguyen and Pierre-Yves Oudeyer. Properties for efficient demonstrations to a socially guided intrinsically motivated learner. In *21st IEEE International Symposium on Robot and Human Interactive Communication*, 2012.

[2] Sao Mai Nguyen and Pierre-Yves Oudeyer. Interactive learning gives the tempo to an intrinsically motivated robot learner. In *IEEE-RAS International Conference on Humanoid Robots*, 2012.

[3] Sao Mai Nguyen and Pierre-Yves Oudeyer. Active choice of teachers, learning strategies and goals for a socially guided intrinsic motivation learner. *Paladyn Journal of Behavioural Robotics*, 3(3):136–146, 2012. ISSN 2080-9778. doi: 10.2478/s13230-013-0110-z. URL http://dx.doi.org/10.2478/s13230-013-0110-z.

[4] Sao Mai Nguyen, Adrien Baranes, and Pierre-Yves Oudeyer. Bootstrapping intrinsically motivated learning with human demonstrations. In *IEEE International Conference on Development and Learning*, Frankfurt, Germany, 2011.