Teaching a Robot with Unlabeled Instructions: The TICS Architecture

JAAMAS Track

Anis Najar
Laboratoire de Neurosciences Cognitives Computationnelles, INSERM U960
Paris, France
anis.najar@ens.fr

Olivier Sigaud
Institute for Intelligent Systems and Robotics, Sorbonne Université, CNRS UMR 7222
Paris, France
sigaud@isir.upmc.fr

Mohamed Chetouani
Institute for Intelligent Systems and Robotics, Sorbonne Université, CNRS UMR 7222
Paris, France
chetouani@isir.upmc.fr

ABSTRACT

In this work, we propose a framework that enables a human to teach a robot a new task by interactively providing it with unlabeled instructions. We ground the meaning of instruction signals in the task-learning process, and use them simultaneously for guiding the latter. We implement our framework as a modular architecture, named TICS (Task-Instruction-Contingency-Shaping) that combines different information sources: a predefined reward function, human evaluative feedback and unlabeled instructions. This approach provides a novel perspective for robotic task learning that lies between Reinforcement Learning and Supervised Learning paradigms. We evaluate our framework both in simulation and with a real robot. The experimental results demonstrate the effectiveness of our framework in accelerating the task-learning process and in reducing the number of required teaching signals.

KEYWORDS

Interactive Machine Learning; Human-Robot Interaction; Reinforcement Learning; Unlabeled Instructions

ACM Reference Format:
Anis Najar, Olivier Sigaud, and Mohamed Chetouani. 2021. Teaching a Robot with Unlabeled Instructions: The TICS Architecture: JAAMAS Track. In *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, May 3–7, 2021*, IFAAMAS, 2 pages.

1 INTRODUCTION

Two complementary approaches are usually considered for task learning in Robotics: autonomous learning [1] and interactive learning [2]. While the main advantage of autonomous learning is the autonomy of the learning process, this approach suffers from several limitations when applied to real-world problems, such as slow convergence and unsafe exploration [1]. By contrast, interactive learning methods overcome these limitations, but come at the cost of human burden during the teaching process, and the cost of predetermining the meaning of teaching signals [3].

In this work [4], we propose a novel framework for robotic task learning that combines the benefits of both autonomous and interactive learning approaches. First, we consider reinforcement learning with a predefined reward function for ensuring the autonomy of the learning process. Second, we consider two types of human-provided teaching signals, evaluative feedback and instructions, for accelerating the learning process. Moreover, we relax the constraint of predetermining the meaning of instruction signals by making the robot incrementally interpret their meaning during the learning process. Our main contribution is to show that instructions can effectively accelerate task learning, even without predetermining their meaning.

We consider interactively provided instruction signals (*e.g.* pointing to the left/to the right) that indicate to the robot which action it has to perform in a given situation (*e.g.* turn left/turn right). Our main idea is to use instruction signals as a means for transferring the information about the optimal action between several task states: all states associated with the same instruction signal collectively contribute to interpreting the meaning of that signal; and in turn, an interpreted signal contributes to learning the optimal action in all task states to which it is associated. This scheme serves as a bootstrapping mechanism that reduces the complexity of the learning process; and constitutes a novel perspective for robotic task learning that lies between Reinforcement Learning and Supervised Learning paradigms. Under this scheme, unlabeled instructions are interpreted by reinforcement learning, and used for labeling task states by supervised learning.

2 THE TICS ARCHITECTURE

We implement our framework as a modular architecture, named TICS (Task-Instruction-Contingency-Shaping), which combines different information sources: a predefined reward function, human evaluative feedback and unlabeled instructions. The general architecture is based on four components: a Task Model (TM), an Instruction Model (IM), a Contingency Model (CM) and a Shaping Component (SC) (Fig. 1).

The Task Model is responsible for learning the task from rewards and/or evaluative feedback, while the Instruction Model is responsible for interpreting instructions. The Contingency Model links task states within TM to instruction signals within IM, by determining which signal has been observed in each state. The role of this model is to minimize the number of interactions with the teacher by recalling the previously provided instructions, and also to make the mapping between states and instructions signals more robust to errors. Finally, the Shaping Component is responsible for combining the outputs of TM and IM for decision-making.

Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), U. Endriss, A. Nowé, F. Dignum, A. Lomuscio (eds.), May 3–7, 2021, Online.

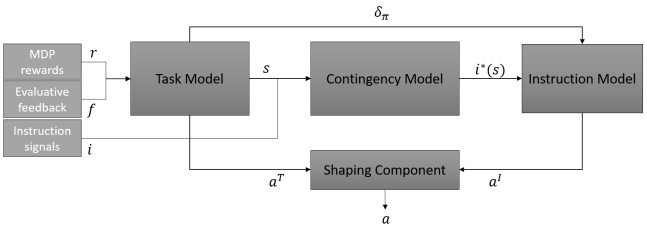
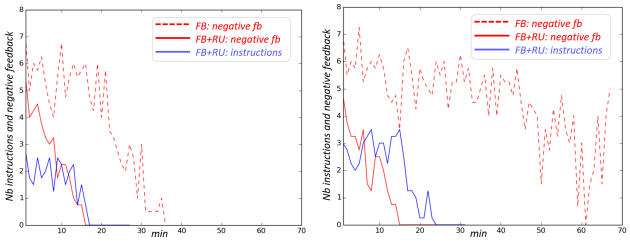


Figure 1: The TICS architecture includes four main components: a Task Model learns the task, a Contingency Model associates task states with instruction signals, an Instruction Model interprets instructions, and a Shaping Component combines the outputs of the Task Model and the Instruction Model for decision-making.



(a) Small state space. **(b) Large state space.**

Figure 2: Number of instructions (blue) and negative feedback (red) over time. FB: using feedback only. FB+RU: using feedback plus unlabeled instructions.

3 RESULTS IN SIMULATION

We first evaluate our framework in simulation on two different problems: object sorting [5] and maze navigation. Simulations allow us to systematically evaluate the performance of our system under different hypotheses about the teaching conditions, and to test its limits under worst case scenarios. For instance, we evaluate the robustness of our framework against various levels of sparse and erroneous teaching signals. The experimental results of our simulations can be summarized as follows:

Ideal case: When teaching signals are correct and not sparse, our framework improves the convergence rate with respect to learning without unlabeled instructions.

Sparse instructions: When learning from evaluative feedback, our framework is robust against all levels of instruction sparsity, and improves the convergence rate with respect to not using unlabeled instructions. However, when learning from a reward function, the existence of multiple possible interpretations can prevent the learning process from converging. This only happens in domains with multiple optimal policies and when instructions are below a certain level of sparsity. When the reward function is combined with evaluative feedback, our framework becomes robust against all levels of instruction sparsity, as feedback enables the teacher to rectify misinterpreted instructions.

Erroneous instructions: Our framework is robust against erroneous instructions and improves the convergence rate, if the probability of receiving erroneous instructions is lower than 0.3.

Sparse feedback: When learning only from evaluative feedback, our framework improves the convergence rate and the robustness of the learning process against feedback sparsity. However, it is still limited to a certain level of sparsity. With a reward function, the learning process becomes robust against all levels of feedback sparsity.

Erroneous feedback: Our framework is robust against erroneous feedback and improves the convergence rate, if the probability of receiving erroneous feedback is lower than 0.5.

Interaction load: In the ideal case, our framework reduces the number of evaluative feedback and the total number of required teaching signals.

4 EXPERIMENT WITH A REAL ROBOT

We also evaluate our framework with a real robot and a real human teacher on the object sorting task[1]. We assess the performance of the TICS architecture when using unlabeled instructions with respect to only using evaluative feedback. In order to assess the scalability of our framework to different task complexities, we contrast two experimental conditions by varying the complexity of the state-space representation.

Figure 2 reports the evolution of the number of provided instructions and negative feedback over time for each condition. In the small state space condition, the baseline model converges after 36 minutes, while our model converges within 17 minutes. In the large state space condition, the baseline model does not completely converge after an hour of training, while our model converges within 24 minutes.

These results are consistent with those obtained in simulation and with the results reported by [5]. They show that our model reduces considerably the number of steps and training time. We also find that our model achieves better performance with less interactions. The robot also explores fewer states and spends less time in each of them, which reflects a more efficient exploration strategy.

5 CONCLUSION

This work presents a novel framework for teaching a robot a new task with unlabeled human instructions. The key idea is to reduce the complexity of a task-learning process through unlabeled instruction signals. These signals are interpreted by the robot, and used simultaneously for accelerating the task-learning process.

REFERENCES

[1] J. Kober, J. A. Bagnell, and J. Peters. Reinforcement Learning in Robotics: A Survey. *Int. J. Rob. Res.*, 32(11):1238–1274, Sept. 2013.
[2] S. Chernova and A. L. Thomaz. Robot learning from human teachers. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 8(3):1–121, 2014.
[3] A. Najar and M. Chetouani. Reinforcement learning with human advice: A survey. arXiv preprint arXiv:200511016, 2020.
[4] A. Najar, O. Sigaud, and M. Chetouani. Interactively shaping robot behaviour with unlabeled human instructions. *Autonomous Agents and Multi-Agent Systems* 34, 35, 2020.
[5] H. B. Suay and S. Chernova. Effect of human guidance and state space size on Interactive Reinforcement Learning. In *2011 RO-MAN*, pages 1–6, July 2011.

[1] https://youtu.be/TK9SwFedtUc