Learning Multiple Tasks with Non-stationary Interdependencies in Autonomous Robots

Extended Abstract

Alejandro Romero
Richard J. Duro
Integrated Group for Engineering Research, CITIC research center, Universidade da Coruña
A Coruña, Spain
{alejandro.romero.montero,richard.duro}@udc.es

Gianluca Baldassarre
Vieri Giuliano Santucci
Institute of Cognitive Sciences and Technologies (ISTC), National Research Council of Italy (CNR)
Rome, Italy
{gianluca.baldassarre,vieri.santucci}@istc.cnr.it

ABSTRACT

An important challenge in the field of autonomous open-ended learning is the autonomous learning of interdependent tasks, and in particular when such interdependencies are non-stationary, so that the robot has to modify the acquired knowledge to properly sequence goals that constitute preconditions for other ones. This work proposes a hierarchical robotic architecture to address this type of scenarios, allowing for the autonomous learning of both the skills necessary to achieve the multiple goals, and of the sequences reflecting the relations between them. Moreover, our system is endowed with a mechanism that, on the basis of self-estimated competence over goal achievement, is able to self-tune the exploration-exploitation balance to cope with the non-stationarity of the environment. The architecture is tested using an UR5e robot operating in a scenario where it should autonomously learn to accomplish various manipulation tasks.

KEYWORDS

Developmental Robotics; Machine Learning for Robot Control; Cognitive Control Architectures

ACM Reference Format:
Alejandro Romero, Richard J. Duro, Gianluca Baldassarre, and Vieri Giuliano Santucci. 2023. Learning Multiple Tasks with Non-stationary Interdependencies in Autonomous Robots: Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

1 INTRODUCTION

Autonomy is a crucial feature for the development of robotic systems that can be deployed in real-world scenarios, where robots must adapt to situations not foreseen at design-time and learn new skills to achieve their goals eventually handling unexpected changes in the environment. Amongst other approaches [7–9], intrinsically motivated open-ended learning [1, 10] leverages on "curiosity" and self-generated motivational signals [15] to build agents (often implemented in the reinforcement learning framework [5]) that can autonomously select their own goals and acquire the behaviours necessary to achieve them [3, 4, 6, 16]. However, in spite of the considerable advances that this line of research has managed to make towards increasingly autonomous systems, the majority of works in the field [2, 16] focus on the autonomous learning of multiple independent tasks, that is, tasks not related to each other. Even less investigated, especially with real robots, are scenarios in which the interdependencies between tasks are *non-stationary*, thus requiring the robot to modify its behaviour and the representation of its knowledge. In this work we propose an architecture providing a solution for robots operating in such scenarios.

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

2 IMPLEMENTED SOLUTION

Leveraging and extending previous research [11, 13, 14, 17], we propose a hierarchical cognitive architecture that encompasses different mechanisms and functions to autonomously: (1) select goals on the basis of IMs; (2) select and learn sequences of interdependent (sub-)goals needed to achieve the desired one (autonomous curriculum learning); (3) cope with non-stationary interdependencies between goals. The system is composed of 3 main layers, each working at a different temporal level (epochs, trials and time-steps):

- The **Goal-Selector**, implemented as an N-armed bandit, determines the goal to pursue in the current **epoch** (composed of n trials) according to a *softmax* selection rule based on the current values of the goals. These values are updated through a standard exponential moving average (EMA) of the competence improvements ΔC^g obtained when training on each goal. In particular, ΔC^g is calculated as the difference between two averages of predictions (CP), each one over a period PT of 10 attempts related to the same goal:

$$\Delta C^g = \frac{\sum_{i=t-(PT-1)}^{t} |CP_i|}{PT} - \frac{\sum_{i=t-(2PT-1)}^{t-PT} |CP_i|}{PT} \quad (1)$$

  where CP_i is the difference between the prediction generated at the beginning of the trial (expected probability of achieving g) and the actual result of the robot's attempt (1 for success, 0 otherwise).
- The **Sub-Goal Selector**, implemented as a Q-Learning algorithm, at each **trial** receives as inputs the selected goal g and the current state of the environment with respect to the goals, i.e., a binary vector stating if a goal has been achieved during the current epoch. Given this information, the sub-goal to be pursued is selected according to a *softmax* selection rule based on the Q-values of the sub-goals, depending on goal specific reinforcements r^g obtained when achieving g. To cope with the non-stationarity

Figure 1: Experimental setup and relations between goals.

Goal name	Description
Goal 1	Orange button pressed
Goal 2	Green button pressed
Goal 3	Blue button pressed
Goal 4	Red cylinder grabbed
Goal 5	Blue cylinder grabbed
Goal 6	Blue cylinder in square box

Table 1: Achievable goals/sub-goals.

of goal interdependencies, the sub-goal selector modifies its exploratory attitude on the basis of the competence of the system in achieving the selected goals. In particular, the temperature of the *softmax* determining the noise level for the selection of the sub-goals is proportional to $1 - \Delta C^g$.

- The **Experts** are represented as utility models and implemented as neural-networks value functions (see [12]), each one associated with a specific goal. The network receives as input the state of the robot (sensors values) and provides as output the expected utility (probability of reaching the associated goal modulated by the achieved utility). Thus, at each **time step**, the robot generates a series of candidate actions and uses the value function associated with the goal to choose the most appropriate one.

3 EXPERIMENTAL RESULTS

To test the proposed architecture we created a robotic scenario that includes a UR5e robot placed in front of a table where there are several cylinders, boxes and buttons (Fig. 1 left). The boxes are opened by pressing the buttons, which allows us to create interdependencies between objects and make them change over time (see Fig. 1 right). Thus, for one goal to be accessible, it is necessary to previously reach another goal or goals. The different goals and sub-goals the robot can achieve are defined in Table 1.

In the experiments we controlled the direction of movement and the height of the arm. The perception of the robot (input to utility models) at each instant of time was: $P(t) = (d_1, \ldots, d_n, s_1, \ldots, s_m)$ where d_j are the relative distances between the objects and the robot end-effector, and s_i are the states of the different buttons.

The experiment was run 20 times and for 2,000 epochs each run. Each epoch finished after a maximum of 8 trials or when the main goal was achieved. After that, the scenario was reset. Each trial ended when the robot reached the selected sub-goal or after a timeout of 70 time steps. Interdependencies between objects (see Fig. 1 right) changed every 1,000 epochs.

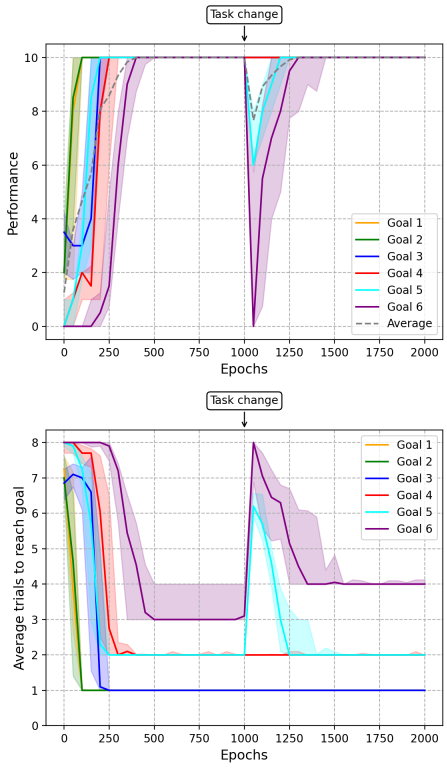

Figure 2: (Top) performance in achieving the different goals of the scenario. (Bottom) trials needed to achieve the goals.

To evaluate the learning of the system, Fig. 2 (top) shows the performance of the robot in achieving each goal. Thus, we can see if it was able to learn the skills and the interdependencies necessary to achieve the goals. In addition, to evaluate the efficiency of the robot in carrying out the tasks, that is, that it only performs the optimal number of sub-goals selections to reach the goal, Fig. 2 (bottom) shows the sub-goal selections (trials) needed to achieve the goal set by the goal selector.

The results show the robot can efficiently adapt to the change of interdependencies and achieve 100% performance in all of them. A video illustrating the robot's performance is available at https://github.com/alejandro-romero/AAMAS_2023.

4 CONCLUSIONS

The innovation of the presented architecture is two-fold: on the one hand, the de-coupling into two different layers of the mechanisms related to autonomous goal selection based on IMs, and the selection and learning of the sequences of sub-goals necessary to achieve the desired one; on the other hand, the capability of the system to face non-stationary scenarios where interdependencies between goals can change, based on a mechanism that employs a measure of competence to autonomously regulate the exploration-exploitation balance. The experimental results have shown the effectiveness of the proposed architecture and paved the way to its application to scenarios involving compound objects and unstructured conditions that can change in time.

ACKNOWLEDGMENTS

This work was partially supported by the MCIU of Spain/FEDER (grant RTI2018-101114-B-I00), Xunta de Galicia (EDC431C-2021/39), Centro de Investigación de Galicia "CITIC" (ED431G 2019/01), and partially by the European Union's Horizon 2020 Research and Innovation Programme under GA no 713010 ('GOAL-Robots – Goal-based Open-ended Autonomous Learning Robots') and GA 945539 ('Human Brain Project, BP SGA3'), and partially by Horizon Europe, GA 101070381 ('PILLAR-Robots - Purposeful Intrinsically motivated Lifelong Learning Autonomous Robots').

REFERENCES

[1] Gianluca Baldassarre and Marco Mirolli. 2013. *Intrinsically motivated learning in natural and artificial systems.* Springer.

[2] Adrien Baranes and Pierre-Yves Oudeyer. 2013. Active learning of inverse models with intrinsically motivated goal exploration in robots. *Robotics and Autonomous Systems* 61, 1 (2013), 49–73.

[3] Andrew G Barto, Satinder Singh, and Nuttapong Chentanez. 2004. Intrinsically motivated learning of hierarchical collections of skills. In *Proceedings of the 3rd International Conference on Development and Learning*. 112–19.

[4] Cédric Colas, Pierre Fournier, Mohamed Chetouani, Olivier Sigaud, and Pierre-Yves Oudeyer. 2019. Curious: intrinsically motivated modular multi-goal reinforcement learning. In *International conference on machine learning*. PMLR, 1331–1340.

[5] Cédric Colas, Tristan Karch, Olivier Sigaud, and Pierre-Yves Oudeyer. 2022. Autotelic agents with intrinsically motivated goal-conditioned reinforcement learning: a short survey. *Journal of Artificial Intelligence Research* 74 (2022), 1159–1199.

[6] Muhammad Burhan Hafez and Stefan Wermter. 2021. Behavior Self-Organization Supports Task Inference for Continual Robot Learning. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 6739–6746.

[7] Alexander S Klyubin, Daniel Polani, and Chrystopher L Nehaniv. 2008. Keep your options open: An information-based driving principle for sensorimotor systems. *PloS one* 3, 12 (2008), e4018.

[8] John Lones, Matthew Lewis, and Lola Cañamero. 2016. From sensorimotor experiences to cognitive development: Investigating the influence of experiential diversity on the development of an epigenetic robot. *Frontiers in Robotics and AI* 3 (2016), 44.

[9] Marlos C Machado, Marc G Bellemare, and Michael Bowling. 2017. A laplacian framework for option discovery in reinforcement learning. In *International Conference on Machine Learning*. PMLR, 2295–2304.

[10] Pierre-Yves Oudeyer, Frdric Kaplan, and Verena V Hafner. 2007. Intrinsic motivation systems for autonomous mental development. *IEEE transactions on evolutionary computation* 11, 2 (2007), 265–286.

[11] Alejandro Romero, Gianluca Baldassarre, Richard J Duro, and Vieri Giuliano Santucci. 2021. Analysing autonomous open-ended learning of skills with different interdependent subgoals in robots. In *2021 20th International Conference on Advanced Robotics (ICAR)*. IEEE, 646–651.

[12] A. Romero, F. Bellas, A. Prieto, and R.J. Duro. 2018. Utility Model Re-description within a Motivational System for Cognitive Robotics. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2324–2329. https://doi.org/10.1109/IROS.2018.8593799

[13] Alejandro Romero, Abraham Prieto, Francisco Bellas, and Richard J Duro. 2019. Simplifying the creation and management of utility models in continuous domains for cognitive robotics. *Neurocomputing* 353 (2019), 106–118.

[14] Vieri Giuliano Santucci, Gianluca Baldassarre, and Emilio Cartoni. 2019. Autonomous reinforcement learning of multiple interrelated tasks. In *2019 Joint IEEE 9th international conference on development and learning and epigenetic robotics (ICDL-EpiRob)*. IEEE, 221–227.

[15] Vieri Giuliano Santucci, Gianluca Baldassarre, and Marco Mirolli. 2013. Which is the best intrinsic motivation signal for learning multiple skills? *Frontiers in neurorobotics* 7 (2013), 22.

[16] Vieri Giuliano Santucci, Gianluca Baldassarre, and Marco Mirolli. 2016. Grail: a goal-discovering robotic architecture for intrinsically-motivated learning. *IEEE Transactions on Cognitive and Developmental Systems* 8, 3 (2016), 214–231.

[17] Vieri Giuliano Santucci, Davide Montella, and Gianluca Baldassarre. 2022. C-GRAIL: Autonomous reinforcement learning of multiple, context-dependent goals. *IEEE Transactions on Cognitive and Developmental Systems* (2022).