BY-NC-ND

Mechanisms for Autonomous Sub-goal Discovery in Lifelong Robotic Learning

Emanuel Fallas Hernández, Sergio Martínez Alonso, José A. Becerra Permuy, Alejandro Romero, and Richard J. Duro

Grupo Integrado de Ingeniería, Centro de Investigación CITIC, Universidade da Coruña, 15071 A Coruña, Spain
Correspondence: emanuel.fallas@udc.es

DOI: *https://doi.org/10.17979/spu.23.c53*

Abstract: Lifelong learning in robotics seeks the continuous acquisition and reuse of knowledge to autonomously master complex tasks. A key challenge is discovering sub-goals without relying on explicit intermediate rewards, enabling robots to decompose tasks and transfer skills across domains. We propose a dual mechanism for sub-goal discovery. The first is a top-down strategy that builds hierarchical sub-goal chains from general goals via intrinsic motivations. The second is a bottom-up approach that uncovers latent links between previously learned goals and perceptual classes. Implemented in the e-MDB cognitive architecture, our method was tested in both simulation and a real-world robotic manipulation task. Results show efficient sub-goal generation, transfer, and generalization.

1 Introduction

Lifelong learning in robots seeks to move beyond machines that can only perform tasks designed for them in advance. Instead, it focuses on creating robots that can keep learning throughout their entire existence, adapting to new situations and building on what they have already learned Romero et al. (2023). This approach is especially important as robots are expected to operate in environments that change in unpredictable ways, where pre-programmed solutions are not enough.

To implement such systems, cognitive architectures emerge as the foundation for intelligent behavior. This is achieved by the integration of perceptual information and learning processes that aim towards the fulfillment of robot's goals. As cognitive architecture development has progressed, they have evolved from rule-based architectures Laird et al. (1987); Lebiere and Anderson (1993); Varma and Just (2006) to connectionist architectures that leverage machine learning algorithms for continuous adaptation and learning Becerra et al. (2021); Kotseruba and Tsotsos (2020); Romero et al. (2022); Weng (2004).

A fundamental characteristic of real environments is that tasks are openly defined and it might not be possible to provide an adequate measurement of progress for each step taken by the robot. This provides a significant challenge for learning algorithms that rely on a explicit reward value for each state, as is the case for many reinforcement learning algorithms Kulkarni et al. (2016). For this reason, mechanisms for the generation of sub-goals are key for achieving lifelong learning capabilities for robots. Additionally, having a representation for each of those intermediate steps allow for reuse of learned skills in different tasks and also aid with the explainability of the autonomous behavior.

Instead of treating every sensor reading as unique, perceptual classes group together states that share the same functional "meaning" Feldman (1992). For example, many different camera images may correspond to the presence of a chair, and all of these sensory states can be assigned to the perceptual class *chair*. Formally, a perceptual class $\hat{s}$ is defined as as a subset of the robot's state space S, such that all states $s \in \hat{s}$ represent situations with equivalent relevance for decision-making. By clustering states in this way, perceptual classes provide the robot with stable categories that simplify reasoning, reduce complexity, and allow knowledge to be reused across different tasks.

Using this definition, robot's goals can be represented using perceptual classes. A goal g can be defined as the subset of states in S that the robot tries to reach. In conjunction with goals, other perceptual classes allow the robot to identify important states that are required to reach them. These can be thought as transition states toward goals. By autonomously generating these perceptual classes through the interaction with the environment, robots are able to generate grounded representations of their knowledge. These allow for simplified decision-making processes as if symbolic representations were used.

In this paper we propose two mechanisms for the discovery of sub-goals: a top-down approach, where sub-goals are hierarchically derived from general goals based on intrinsic motivations Romero et al. (2020), and a bottom-up approach, where sub-goal chains emerge from goals and non-goal perceptual classes that were previously learned in different domains Romero et al. (2022). In the following section we provide an explanation of the mechanisms proposed. Then, in section 3 we describe the experiments that were carried out. In section 4 the main results obtained are analyzed and finally, section 5 provides the conclusions obtained.

2 Proposed Mechanisms

The cognitive architecture used in this work is the epistemic Multilevel Darwinist Brain (e-MDB). e-MDB integrates perception, motivation, memory, and action-selection modules into a unified system, allowing robots to autonomously acquire skills and refine their knowledge over time Bellas et al. (2010); Romero et al. (2023). The architecture provides mechanisms for obtaining and relating perceptual classes and goals Becerra et al. (2021). Therefore, it provides a suitable framework for testing strategies of autonomous sub-goal generation in dynamic environments.

Knowledge in the cognitive architecture is stored by using a Memory System Duro et al. (2019). It is organized by the activation relationships of knowledge nodes in a directed graph. These relationships are represented as C-Nodes (C_n), which relate policies (π) with a World Model (WM), a goal (G) and the prior perceptual state, called P-Node (P_n).

Perceptual classes are used to represent subsets of the robot's state space, either as preconditions to achieve a goal (P-Nodes) or as states that yield a reward (goal nodes). These classes are learned progressively during the robot's lifetime, and therefore, latent relationships must be searched only after the representations have stabilized. This is evaluated through a confidence measure C based on classification history.

To generate useful sub-goals in lifelong learning, two requirements must be met: *generalization*, so that sub-goals capture all states leading to a target goal, and *reusability*, so that knowledge from different domains can be adapted in new contexts. In our approach, perceptual classes in the e-MDB architecture are used to define new goals or link existing ones.

2.1 Top-down sub-goal Discovery

The top-down approach derives sub-goals from existing knowledge of how to reach a final goal. Sub-goals are created from the states associated with learned P-Nodes, this is, the states before reaching the final goal. This process is driven by an internal mechanism based on the effectance drive Romero et al. (2020), which activates once a P-Node surpasses a confidence threshold. When this occurs, a new goal is generated that rewards the robot each time the P-Node is activated, as illustrated in Fig. 1.

Successive executions of this mechanism allows a chain of sub-goal to be created. Therefore,

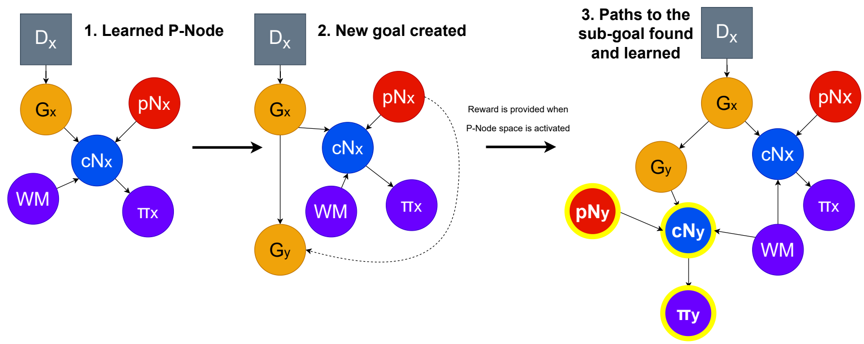

Figure 1: Top-down sub-goal generation.

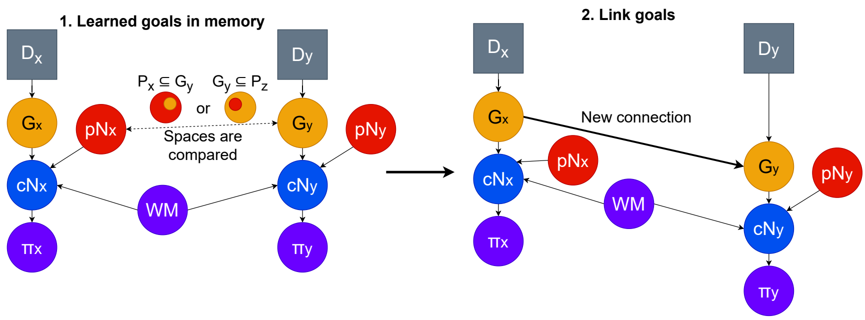

Figure 2: Bottom-up sub-goal generation. G_y is identified as sub-goal of G_x.

the representation of a goal and its sub-goals is made in a step-by-step basis. The explicit representation of each step then aids with the reuse of the obtained knowledge in different contexts.

2.2 Bottom-up sub-goal Discovery

The bottom-up approach reuses accumulated knowledge by analyzing latent relationships between perceptual classes of P-Nodes and goals. If a P-Node is contained within a goal ($P_i \subseteq G_j$) or vice versa ($G_j \subseteq P_i$), the downstream goal can be considered a sub-goal of the upstream goal. This is analogous to a prospection process, where candidate states are evaluated for their utility in reaching other goals. Once a match is found, the activation of upstream goals is cascaded toward sub-goals, as shown in Fig. 2.

To ensure robust behavior, safeguards are included: feedback loops in goal chains are avoided, and activation is attenuated when cascading from a goal to its sub-goals. This ensures final goals to retain priority over intermediate ones.

3 Experimental Setup

The experiment consisted of a fruit classification task, where a TIAGo robot had to sort fruits into accepted or rejected baskets based on their weight (Fig. 3). Fruits were collected from a designated area, weighed on a scale, and classified accordingly. An overhead RGB camera provided the robot with the position and size of the nearest fruit, along with information about the scale and a table button light (the latter being irrelevant to the classification task). The cognitive architecture operated on a ten-dimensional perceptual state that encoded fruit location, grasping status, scale readings, and the state of the button light.

Figure 3: Experimental setup with a TIAGo robot.

Eight policies were available to complete the task, these are described in Table 1. Perceptual classes were modeled using neural networks with three hidden layers (128, 64, and 32 neurons), ReLU activations, and a sigmoid output. Networks were trained online with binary cross entropy loss, the Adam optimizer, and a FIFO buffer of 400 samples to prevent overfitting. Training occurred only when predictions diverged from labels.

Table 1: Policies available to the robot.

Policy	**Description**
Pick fruit	Grasp a fruit from the table
Test fruit	Place fruit on the scale for evaluation
Accept fruit	Place fruit in accepted basket
Discard fruit	Place fruit in rejected basket
Change hands	Transfer fruit between grippers
Place fruit	Leave fruit in the center of the table
Ask nicely	Request more fruits from experimenter
Press button	Toggle table button light

To evaluate sub-goal generation, three experiments were compared: (1) a baseline where rewards were given for every correct action, (2) a setup combining both top-down and bottom-up sub-goal generation, and (3) a classical top-down only approach. A learning curriculum simulated lifelong acquisition of goals and perceptual classes:

1. Exploration phase. The weighing scale was inactive.
2. The robot was tasked with placing fruit in the middle of the table.
3. Exploration phase with the scale enabled.
4. The robot was tasked with classifying fruit.

During the exploratory phases, goals were created using an effectance-based intrinsic motivation Romero et al. (2020). This generated goals related to events in the perception, such as grasping a fruit.

4 Results and Discussion

The three experimental setups were evaluated in a discrete event simulator, with 750 trials per experiment which consisted of 20 iterations each. The fist phase of the curriculum was active for 20 trials, the second up to trial 125, the third from trial 126 to trial 150 and the final phase the remaining trials. Ten runs were executed for each of the methods.

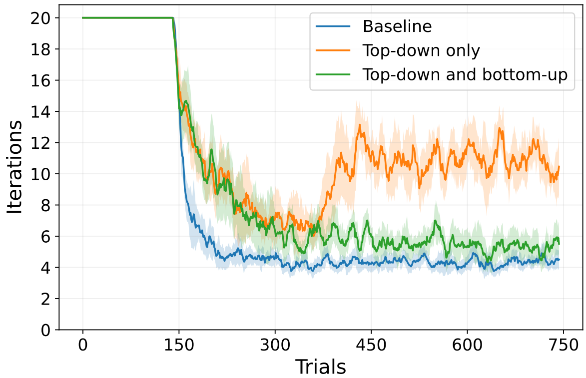

Figure 4: Performance comparison of the different sub-goal generation approaches.

Figure 4 compares performance: while the baseline reaches on average 4 iterations per trial, the top-down only approach stabilizes at 11, and the combined top-down + bottom-up approach achieves close-to-baseline performance with 5 iterations per trial. It should be noted that in the baseline the robot receives step-by-step rewards, whereas in the autonomous approaches rewards are only given upon reaching the final goal.

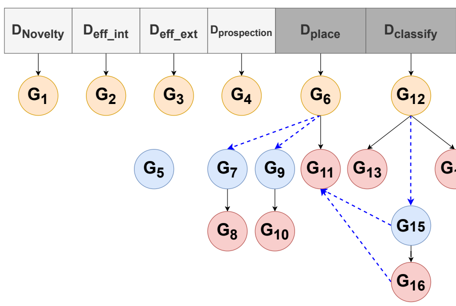

Figure 5: Goal tree after the Top-down + Bottom-up experiment. Light blue: environment-effect goals. Red: top-down sub-goals. Blue arrows: bottom-up links.

Figures 5 and 6 show the final goal trees for the combined and top-down only approaches. In both cases, the graphs reflect the sequence of the learning curriculum: environment-effect goals appear first, followed by task-related goals. The top-down mechanism creates sub-goals from P-Nodes, while the bottom-up mechanism establishes latent links across existing goals, which results in richer structures. Table 2 summarizes the goals discovered in memory for each case.

The results show that the combined approach enables reuse of previously learned goals and perceptual classes, effectively bootstrapping new skills. For example, placing a fruit on the table (G_6, Top-down + Bottom-up) leverages prior grasping goals (G_7, G_9, Top-down + Bottom-up), which later support learning to activate the scale (G_{15}, Top-down + Bottom-up).

Table 2: Comparison of goals in the LTM after the execution of the experiments.

Top-down only		Top-down + Bottom-up	
Goal	**Description**	**Goal**	**Description**
G_1	Novel state achieved	G_1	Novel state achieved
G_2	Sub-goal created for learned P-Node	G_2	Sub-goal created for learned P-Node
G_3	Goal linked to environment effect created	G_3	Goal linked to environment effect created
G_4	Button light turned on	G_4	Sub-goal link created
G_5	Button light turned off	G_5	Button light turned on
G_6	Button light turned on	G_6	Fruit placed on table
G_7	Fruit grasped with left gripper	G_7	Fruit grasped with left gripper
G_8	Fruit grasped with right gripper	G_8	Fruit grasped with right gripper
G_9	Fruit grasped with right gripper	G_9	Fruit grasped with right gripper
G_{10}	Fruit grasped with left gripper	G_{10}	Fruit grasped with left gripper
G_{11}	Weighing scale activated	G_{11}	Fruit grasped on any gripper
G_{12}	Fruit grasped on the same side as the scale	G_{12}	Fruit classified properly
G_{13}	Fruit placed on table	G_{13}	Good fruit tested
G_{14}	Fruit grasped on any gripper	G_{14}	Bad fruit tested
G_{15}	Fruit classified properly	G_{15}	Weighing scale activated
G_{16}	Good fruit tested	G_{16}	Fruit grasped on the same side as the scale
G_{17}	Bad fruit tested		

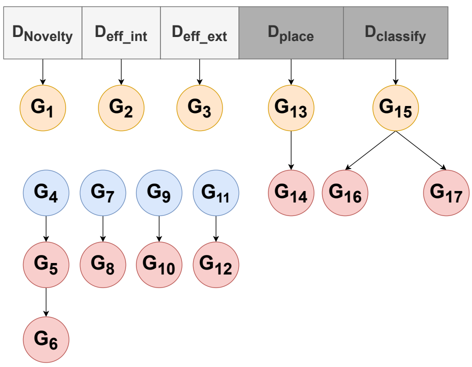

Figure 6: Goal tree after the Top-down only experiment. Light blue: environment-effect goals. Red: top-down sub-goals.

In contrast, the top-down only setup leaves certain sub-goals unlinked. For example, the goal related to activating the weighing scale (G_{11}, Top-down only), is not considered a sub-goal of the classifying task. As this is an environmental effect goal, after the architecture learns how to recreate it, it is no longer pursued. This provokes the important performance drop seen around trial 400 in Figure 4, as there is no bootstrapping effect from completing that goal.

5 Conclusions

This paper addressed the challenge of sub-goal generation in lifelong open-ended learning, where achieving goals often requires a sequence of intermediate steps. A system based on two mechanisms for sub-goal generation was presented. First, a traditional top-down approach that progressively build hierarchies of sub-goals from known goal states. This was complemented by a bottom-up approach that exploits the robot's accumulated knowledge in perceptual classes to discover latent relationships between goals and sub-goals.

Experimental results in a manipulation task showed that our proposal achieves performance close to the optimal baseline, despite only rewarding the final goal, and improves efficiency by more than a factor of two compared to the classical top-down approach alone. These results showed that the mechanisms were able to reuse knowledge obtained in previously learned tasks to improve its performance in a new task.

Future work will focus on developing more advanced algorithms for uncovering latent relationships within the stored perceptual classes stored in memory. So that more complex tasks can be achieved autonomously.

Bibliography

J. A. Becerra, A. Romero, F. Bellas, and R. J. Duro. Motivational engine and long-term memory coupling within a cognitive architecture for lifelong open-ended learning. *Neurocomputing*, 452:341–354, 2021.

F. Bellas, R. J. Duro, A. Faiña, and D. Souto. Multilevel darwinist brain (mdb): Artificial evolution in a cognitive architecture for real robots. *IEEE Transactions on autonomous mental development*, 2(4):340–354, 2010.

R. J. Duro, J. A. Becerra, J. Monroy, and F. Bellas. Perceptual generalization and context in a network memory inspired long-term memory for artificial cognition. *International Journal of Neural Systems*, 29(06):1850053, 2019.

J. Feldman. Constructing perceptual categories. In *CVPR*, pages 244–250, 1992.

I. Kotseruba and J. K. Tsotsos. 40 years of cognitive architectures: core cognitive abilities and practical applications. *Artificial Intelligence Review*, 53(1):17–94, 2020.

T. D. Kulkarni, K. Narasimhan, A. Saeedi, and J. Tenenbaum. Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL *https://proceedings.neurips.cc/paper_files/paper/2016/file/f442d33fa06832082290ad8544a8da27-Paper.pdf*.

J. E. Laird, A. Newell, and P. S. Rosenbloom. Soar: An architecture for general intelligence. *Artificial intelligence*, 33(1):1–64, 1987.

C. Lebiere and J. R. Anderson. A connectionist implementation of the act-r production system. In *Proceedings of the fifteenth annual conference of the Cognitive Science Society*, pages 635–640, 1993.

A. Romero, F. Bellas, J. A. Becerra, and R. J. Duro. Motivation as a tool for designing lifelong learning robots. *Integrated Computer-Aided Engineering*, 27(4):353–372, 2020.

A. Romero, G. Baldassarre, R. J. Duro, and V. G. Santucci. Autonomous learning of multiple curricula with non-stationary interdependencies. In *2022 IEEE International Conference on Development and Learning (ICDL)*, pages 272–279. IEEE, 2022.

A. Romero, F. Bellas, and R. J. Duro. A perspective on lifelong open-ended learning autonomy for robotics through cognitive architectures. *Sensors*, 23(3):1611, 2023.

S. Varma and M. A. Just. 4caps: An adaptive architecture for human information processing. In *AAAI Spring Symposium: Between a Rock and a Hard Place: Cognitive Science Principles Meet AI-Hard Problems*, pages 91–96, 2006.

J. Weng. Developmental robotics: Theory and experiments. *International Journal of Humanoid Robotics*, 1(02):199–236, 2004.