Reinforcement Learning

Paola Rojas Domínguez

***Abstract*—The acquisition of complex motor behaviors in robots, such as standing, balancing, or steady flight, poses significant challenges for reinforcement learning due to the high dimensionality of the state space. Hierarchical reinforcement learning (HRL) offers an effective solution by decomposing global goals into manageable subgoals, enabling both strategic planning and detailed motor control. This report analyzes HRL applied to real robots, highlighting a two-tier architecture where a high-level planner defines intermediate goals and a low-level controller executes precise joint actions, leading to faster and more robust learning compared to planar methods. In parallel, insights from neuroscience, particularly Kenji Doya's studies on neuromodulation and mental simulation, are explored as biological analogies inspiring reward shaping, adaptive control, and predictive modeling in robotics. Finally, these ideas extend to drone training, where hierarchical planning, partial rewards, and safe simulation-based exploration are essential for achieving stable and resilient autonomous flight.**

***Index Terms*—Reinforcement Learning, Hierarchical Reinforcement Learning, Neuromodulation, Drones.**

I. Introduction

THE acquisition of complex motor behaviors in robots constitutes one of the most relevant challenges in the field of reinforcement learning. Tasks such as standing, maintaining balance, or flying stably require exploring high-dimensional state spaces, which makes learning difficult when approached in a flat manner. In this context, hierarchical reinforcement learning (HRL) architectures offer an efficient alternative by breaking down global objectives into intermediate subobjectives, facilitating both local exploration and optimization. In turn, neuroscience research, such as those proposed by Kenji Doya, suggests that the brain employs analogous mechanisms through neuromodulators that regulate motivation, decision-making, and motor control. This integrative view not only allows us to better understand how biological systems solve complex problems, but also to transfer these principles to robot design. This paper analyzes two complementary approaches: the application of HRL in physical robots and neurobiological models of mental control and simulation. It ultimately explores their potential adaptation to drone training, where hierarchical planning, partial feedback, and pre-simulation are essential elements for achieving safe and robust learning.

II. Acquisition of stand-up behavior by a real robot using hierarchical reinforcement learning

The article proposes a hierarchical reinforcement learning (HRL) architecture as a practical solution to address the complexity of robot control in high-dimensional dynamic tasks. Instead of trying to have the robot learn the entire task of standing up in a single step, which would involve exploring an excessively large and unwieldy state space, the authors divide the problem into two levels: a higher level, which works in a reduced space and defines intermediate subgoals, and a lower level, which is responsible for translating these subgoals into physical joint movements. In this way, the system simplifies the overall exploration of the state space and allows the robot to move toward its final standing goal more efficiently.

At the upper level, the authors implement Q-learning to select the appropriate sequence of subgoals that will take the robot from the initial position (lying down) to the desired position (standing). This level operates on a set of simplified variables, such as the angles of the main joints and the position of the center of mass, and receives rewards both for achieving the final objective and for meeting intermediate subgoals (Rsub). The inclusion of partial rewards is essential, as it allows the robot to receive positive feedback even in the early stages of learning, preventing the process from stagnating due to lack of reinforcement until complete success. Thus, the upper level plays the role of strategic planner, defining a global path in terms of intermediate postures.

At the lower level, a continuous actor-critic method is used that allows the robot to learn to execute each subgoal specified by the upper level. Here, all the dynamic variables of the system, such as joint angles and velocities, are considered, and appropriate torques for the joints are generated. This approach divides the challenge into multiple local "microcontrollers," each specialized in achieving a specific intermediate posture. The combination of simplified global exploration at the top level and detailed local optimization at the bottom level results in much faster and more robust learning than using flat reinforcement learning. In experiments, the robot was able to learn to stand up in hundreds of trials in simulation and in a few hundred in the real robot, demonstrating that HRL is a powerful technique for addressing complex motor control tasks in robotics. [1]

III. Kenji Doya: Neuromodulation of inference and control in the cortical circuits

Professor Kenji Doya proposes that the brain can be understood as a biological reinforcement learning system. In this framework, neuromodulators, such as dopamine, serotonin, norepinephrine, and acetylcholine, act as chemical signals that guide how neural circuits process rewards, punishments, and future predictions. For example, dopamine is closely related to reward prediction, while serotonin can modulate patience or risk control. In this way, the brain not only reacts to immediate rewards but also anticipates and simulates future scenarios, adjusting behavior based on past experiences. This approach integrates neurobiology with reinforcement learning, showing how cortical circuits implement principles similar

to computational algorithms, but in a flexible and adaptive manner.

In his experiments, Doya combines biological studies with applications in robotics, highlighting the usefulness of reinforcement learning as a bridge between the two disciplines. In the case of robots, defining clear reward functions—for example, maintaining balance or avoiding falls—enables the machine to learn complex motor skills after multiple attempts, mimicking the brain's trial-and-error processes. However, Doya emphasizes that defining the reward function is critical: excessive punishment can lead the robot (or organism) to avoid potentially useful actions, while an appropriate balance of positive reinforcement allows it to explore more options and develop more sophisticated behaviors. This direct relationship between reward design in RL and motivational regulation in the brain reveals profound parallels in how both systems learn.

Finally, Doya introduces the idea of mental simulation as a fundamental mechanism of the brain, comparable to an additional layer in RL systems. The human brain not only responds to present rewards but is also capable of creating internal representations and predicting hypothetical scenarios to make better decisions before acting, which constitutes a key evolutionary advantage. This ability, associated with areas such as the prefrontal cortex, suggests that the brain combines unsupervised learning, memory, and prediction with reward signals, building a flexible system of control and inference. Studies using techniques such as optogenetics reinforce this view by showing in real time how different circuits respond to neuromodulation. Overall, Doya's work demonstrates that the brain engages in a type of biological reinforcement learning, in which the interaction between rewards, punishments, and mental simulation gives rise to intelligent and adaptive behaviors. [2]

IV. HRL Ideas for a Drone

1) Define the complex task (equivalent to stand-up)
   - Stably take off from the ground without tipping over.
   - Learn to stabilize in stationary flight (hover).
   - Safely transition from hover to hover.
   - Regain stability after disturbances (wind, minor collisions).
2) Higher level (subgoal glider, like in the brain)
   - Simplified subgoals that represent key states:
     - Start engines and lift a few centimeters.
     - Reach desired height (example: 1 meter).
     - Maintain a stable position for N seconds.
     - Move in direction X without losing height.
   - Inspiration from neuroscience: the "higher level" would be like the prefrontal cortex → plan sequences based on partial rewards.
3) Lower level (fine control, reflexes)
   - Control of pitch, roll, yaw, and propeller thrust angles.
   - Fast dynamic correction in the face of small disturbances (wind noise, turbulence).
   - Continuous Actor-Critic: The "actor" adjusts motor power; the "critic" evaluates whether the drone is approaching the sub-target.
   - Analogy to the brain: it would be like the cerebellum → responsible for precision and motor coordination.
4) Rewards and punishments (artificial dopamine)
   - Large reward: reaching and maintaining stable flight.
   - Partial rewards: climbing a certain height, maintaining orientation for X seconds.
   - Punishments: losing stability, falling, crashing.
   - Inspiration from Doya's video: if there is too much punishment, the drone might stop trying to take off → balance is key.
5) Mental Simulation (Safe Exploration)
   - Train first in simulators like Gazebo, Isaac Sim, or AirSim, where the drone "imagines" scenarios before testing on real hardware.
   - Inspiration: The brain performs mental simulation → the drone predicts the consequences of actions before executing them.
6) Possible Training Scenarios
   - Hierarchical Takeoff: First learns to fire the motors, then lift off, then stabilize.
   - Robust Hover: Maintain a stable altitude under different wind conditions.
   - Navigated Flight: Use HRL to divide the task into: reach a waypoint → stabilize → continue.
   - Recovery: If control is lost, learn emergency subgoals (e.g., stabilize before falling).
7) Biological Inspiration
   - Dopamine: Reward for stability and altitude reached.
   - Serotonin: risk control, avoidance of dangerous maneuvers.
   - Norepinephrine: rapid reaction to turbulence.
   - Prefrontal cortex: hierarchical flight plan.
   - Cerebellum: fine motor control.
8) Technical Challenges
   - Define subgoals that are realistic and achievable (neither too easy nor impossible).
   - Design reward functions that prevent unwanted behaviors (e.g., a drone that "prefers" not to move to avoid risk).
   - Transfer what has been learned from simulation to the real drone without losing stability (sim-to-real gap).
   - Ensure safety in real tests (emergency stop systems).

V. Conclusion

In conclusion, hierarchical reinforcement learning (HRL) is presented as an effective tool for addressing highly complex robotic tasks by combining strategic planning at higher levels with precise control at lower levels. Parallels with neuroscience, especially Kenji Doya's ideas on neuromodulation and mental simulation, reinforce the validity of this approach

by showing how biological systems solve similar problems. Applying these principles to the case of drones offers a glimpse of more robust and safer solutions, in which the appropriate definition of rewards, subgoals, and simulation scenarios are key to bridging the gap between virtual learning and real-world execution.

REFERENCES

[1] Doya, K., & Morimoto, J. (2001). Acquisition of stand-up behavior by a real robot using hierarchical reinforcement learning. *ELSEVIER*.

[2] snac. (2022, 21 marzo). *Kenji Doya: Neuromodulation of inference and control in the cortical circuits* [Vídeo]. YouTube. https://www.youtube.com/watch?v=WLAPScS-4ao

[3] Xu, B., Gao, F., Yu, C., Zhang, R., Wu, Y., & Wang, Y. (2023). *OmniDrones: An Efficient and Flexible Platform for Reinforcement Learning in Drone Control.*