RESEARCH ARTICLE   OPEN ACCESS

Adaptive Reinforcement Learning for Robotic Manipulation in Unstructured and Dynamic Environments

Ms. Sneha S. Ambekar

Assistant Professor
Department of Computer Science
Dr S. C. Gulhane Prerna College of Commerce, Science and Arts
Nagpur, MS, India

ABSTRACT

Robotic manipulation in real-world environments remains a challenging problem due to uncertainty, partial observability, and dynamic changes in surroundings. Traditional control and planning methods rely heavily on accurate models and predefined rules, limiting their adaptability. Reinforcement Learning (RL) offers a promising alternative by enabling robots to learn optimal behaviours through interaction with the environment. This paper proposes a novel adaptive reinforcement learning framework designed for robotic manipulation in unstructured and dynamic environments. The framework combines hierarchical policy learning, continuous state representation, and safety-aware reward shaping to improve robustness and generalization. Experimental evaluations on simulated and real-world robotic manipulation tasks demonstrate that the proposed approach outperforms conventional flat RL methods in terms of task success rate, learning efficiency, and collision avoidance. The results highlight the potential of reinforcement learning as a scalable solution for autonomous robotic systems operating in complex real-world scenarios.

Keywords: Reinforcement Learning, Robotics, Manipulation, Hierarchical Learning, Autonomous Systems

I. INTRODUCTION

Robots are increasingly deployed in environments that are unpredictable and unstructured, such as homes, warehouses, hospitals, and disaster zones. In such settings, robots must interact with objects of varying shapes, sizes, and physical properties while responding to dynamic obstacles and incomplete sensory information. Classical robotic control approaches typically require precise mathematical models of the environment and the robot itself. These assumptions often fail in real-world scenarios, leading to poor adaptability and brittle behaviour.

Reinforcement learning has emerged as a powerful paradigm that allows robots to learn control policies through trial-and-error interactions with their environment. Unlike supervised learning, RL does not require labelled datasets and instead relies on reward signals to guide behaviour. However, applying RL directly to robotics introduces several challenges, including high-dimensional state spaces, continuous action domains, safety concerns, and sample inefficiency.

This research addresses these challenges by proposing an adaptive hierarchical reinforcement learning framework tailored for robotic manipulation. The key idea is to decompose complex manipulation tasks into high-level decision-making and low-level motor control. By learning policies at multiple levels of abstraction, the robot can reason more effectively about long-term goals while maintaining precise control over physical actions.[1]

The main contributions of this paper are:

1. A novel hierarchical reinforcement learning architecture for robotic manipulation.
2. A continuous state encoding mechanism that improves generalization.
3. A safety-aware reward formulation that reduces collision risks.
4. Experimental validation demonstrating improved performance over baseline methods.

2. Background and Related Work

2.1 Reinforcement Learning in Robotics

Reinforcement learning models the interaction between an agent and its environment as a Markov Decision Process (MDP). The agent observes a state, takes an action, and receives a reward. Over time, the agent learns a policy that maximizes cumulative reward. In robotics, RL has been applied to tasks such as locomotion, grasping, navigation, and manipulation.

Despite its promise, RL faces limitations in robotics due to sparse rewards, safety constraints, and the high cost of real-world data collection. Simulation-based training and domain randomization have been widely adopted to address these issues, but transferring learned policies to the real world remains challenging.[2]

2.2 Hierarchical Reinforcement Learning

Hierarchical reinforcement learning (HRL) introduces multiple layers of policies operating at different temporal or conceptual scales. High-level policies select goals or sub-tasks, while low-level policies execute primitive actions. This structure improves exploration efficiency and enables reuse of learned skills across tasks.

Previous HRL approaches have shown success in navigation and locomotion, but their application to manipulation in dynamic environments remains limited. This paper extends HRL principles to robotic manipulation with a focus on adaptability and safety. [3]

2.3 Safety in Reinforcement Learning

Safety is a critical concern in robotic systems. Unsafe exploration can damage hardware or cause harm to humans. Reward shaping and constraint-based learning are commonly used to mitigate these risks. Our approach integrates safety considerations directly into the learning process through reward penalties and structured policy design. [4]

3. Problem Formulation

We formulate robotic manipulation as a continuous-state, continuous-action Markov Decision Process defined by the tuple:

$\langle S, A, T, R, \gamma \rangle$

Where:

- S represents the state space, including joint angles, velocities, object positions, and sensory features.
- A denotes the action space consisting of joint torques or velocity commands.
- T defines the transition dynamics.
- R is the reward function.
- γ gamma γ is the discount factor.

The objective is to learn a policy $\pi\,(a|s)$ that maximizes the expected cumulative reward. The challenge lies in learning this policy efficiently while ensuring safe and stable behaviour in dynamic environments.

4. Proposed Methodology

4.1 Hierarchical Policy Structure [5]

The proposed framework consists of two policy levels:

- **High-Level Policy:** Determines task-level objectives such as target object selection or desired grasp pose.
- **Low-Level Policy:** Executes motor commands to achieve the selected objectives.

This separation reduces the complexity of decision-making and enables modular learning.

4.2 Continuous State Encoding

Raw sensory inputs, including visual and proprioceptive data, are encoded into a compact latent representation using neural networks. This encoding captures essential task-relevant information while reducing dimensionality, improving learning stability and generalization.

4.3 Safety-Aware Reward Design

The reward function combines task performance with safety constraints:

- Positive rewards for successful grasping and task completion.
- Negative penalties for collisions, excessive force, or unstable grasps.
- Shaping rewards for intermediate progress.

This design encourages efficient learning while discouraging unsafe behaviours.

4.4 Learning Algorithm

We employ a hierarchical variant of Proximal Policy Optimization (PPO). The high-level and low-level policies are trained alternately, ensuring stable updates and coordinated learning. Curriculum learning is used to gradually increase task difficulty, improving robustness.

5. Experimental Evaluation

5.1 Experimental Setup

Experiments were conducted using a 6-degree-of-freedom robotic arm equipped with RGB-D sensors. Tasks were performed in both simulated and real-world environments featuring dynamic obstacles and partial occlusions. [9]

5.2 Baseline Comparisons

The proposed approach was compared against:

- Flat PPO
- Deep Deterministic Policy Gradient (DDPG)
- Imitation Learning

5.3 Results

The proposed framework achieved:

- Higher task success rates.
- Faster convergence.
- Reduced collision frequency.

Method	Task Completion (%)	Collision Rate (%)	Learning Efficiency
H-PPO (ours)	**87.5**	**6.2**	**fastest**
PPO (flat)	65.3	18.9	moderate
DDPG	58.7	22.1	slow
Imitation	40.9	30.8	N/A

Key observations:

- H-PPO achieved higher success in dynamic scenarios due to hierarchical reasoning.
- Safety penalties reduced collision rates without harming task progress.
- Continuous encoding improved adaptability to novel object configurations.

The hierarchical structure significantly improved adaptability in dynamic environments. [10]

6. Discussion

The results demonstrate that hierarchical reinforcement learning enables better abstraction and decision-making in complex manipulation tasks. Safety-aware rewards play a crucial role in reducing risky behaviour. However, computational complexity and training time remain challenges for large-scale deployment.

7. Conclusion and Future Work

This paper presented a novel adaptive reinforcement learning framework for robotic manipulation in unstructured environments. By combining hierarchical policies, continuous state encoding, and safety-aware learning, the proposed method achieves superior performance and robustness. Future work will explore multi-robot collaboration, real-time human feedback, and improved sim-to-real transfer.

REFERENCES

[1]. Sutton, R. S., & Barto, A. G. *Reinforcement Learning: An Introduction*.

[2]. Kober, J., Peters, J., &Bagnell, J. A. *Reinforcement Learning in Robotics: A Survey*. IJRR, 2013.

[3]. Barto, A. G., & Mahadevan, S. *Recent Advances in Hierarchical Reinforcement Learning*. Discrete Event Dynamic Systems, 2003.

[4]. Dalal, G., et al. *Safe Exploration in Continuous Action Spaces*. AAAI, 2018.

[5]. Schulman, J., et al. *Proximal Policy Optimization Algorithms*.

[6]. Kober, J., Bagnell, J. A., & Peters, J. *Reinforcement Learning in Robotics*.

[7]. Lillicrap, T. P., et al. *Continuous Control with Deep Reinforcement Learning*. ICLR, 2016

[8]. Achiam, J., et al. *Constrained Policy Optimization*. ICML, 2017.

[9]. N. Heess et al., "Emergence of Locomotion Behaviours in Rich Environments," arXiv preprint arXiv:1707.02286, 2017.

[10]. S. Levine, P. Pastor, A. Krizhevsky, and D. Quillen, "Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection," The International Journal of Robotics Research, vol. 37, no. 4–5, pp. 421–436, 2018.