RADemics

Policy Optimization and Actor-Critic Methods in Complex Decision-Making Frameworks

Ishrath, S. Devi,
VISVESVARAYA TECHNOLOGICAL UNIVERSITY,
Velalar College of Engineering and Technology

2. Policy Optimization and Actor-Critic Methods in Complex Decision-Making Frameworks

[1]Ishrath, Assistant Professor, Department of Mathematics, HKBK College of Engineering, Visvesvaraya Technological University, Bangalore 560045, Karnataka, India, khushi.unnisa@gmail.com

[2]S. Devi, Associate Professor, Department of Physics, Velalar College of Engineering and Technology, Erode – 638012, Tamil Nadu, India. sdeviphysics@gmail.com

Abstract

This book chapter explores the integration of Policy Optimization and Actor-Critic methods within complex decision-making frameworks, highlighting their critical role in enhancing autonomous systems and real-time decision processes. It delves into the theoretical foundations of these techniques, emphasizing their applications in dynamic environments with high-dimensional state and action spaces. By examining recent innovations and real-world barriers, the chapter provides a comprehensive understanding of the challenges and opportunities associated with scaling these methods. Special attention was given to emerging advancements in sample efficiency, robustness to uncertainty, and multi-agent interactions, essential for practical deployment in sectors such as robotics, healthcare, and finance. It explores future directions, including the potential for hybrid approaches and improved generalization across domains. This chapter contributes valuable insights into the ongoing evolution of reinforcement learning, offering a roadmap for researchers and practitioners working on advanced decision-making systems.

Keywords:

Policy Optimization, Actor-Critic, Reinforcement Learning, Decision-Making, Multi-Agent Systems, Real-Time Decision Processes.

Introduction

The growing complexity of decision-making environments in autonomous systems has necessitated the development of advanced techniques capable of optimizing policies in dynamic and uncertain contexts [1-3]. At the core of these advancements lies Policy Optimization and Actor-Critic methods, which have garnered significant attention for their ability to handle large-scale, high-dimensional decision-making problems [4,5]. These approaches offer substantial benefits in applications ranging from robotics to healthcare, where systems must learn to adapt to ever-changing environments while maximizing long-term rewards [6-9]. The integration of these methods into complex decision-making frameworks allows for more efficient and robust decision processes, crucial for real-time operations in real-world settings [10].

Policy Optimization techniques focus on directly improving the policy by optimizing a performance objective [11]. By using algorithms such as Proximal Policy Optimization (PPO) or Trust Region Policy Optimization (TRPO), systems can learn optimal actions that maximize rewards in environments with uncertain dynamics [12,13]. These techniques are highly effective in environments where traditional value-based methodsstruggle due to high-dimensional state spaces and the need for continuous action selections [14]. The key advantage of policy optimization was its ability to directly handle continuous action spaces, making it particularly suited for applications like robotics and autonomous vehicles, where precise control was necessary [15,16].

Actor-Critic methods, a hybrid of value-based and policy-based methods, further enhance the performance of decision-making systems [17-19]. The Actor component suggests actions based on the current policy, while the Critic evaluates the action's effectiveness by estimating the value function [20]. This dual approach allows Actor-Critic methods to combine the advantages of both approaches, offering improved sample efficiency and faster convergence [21]. Actor-Critic algorithms, such as Advantage Actor-Critic (A2C) and Deep Deterministic Policy Gradient (DDPG), have been successfully applied in a variety of real-world applications, demonstrating their ability to tackle complex, high-dimensional tasks such as game playing, robotics, and autonomous navigation [22,23].

One of the primary hurdles was the computational resources required to train and deploy these models, particularly when operating in environments with vast state and action spaces [24]. Scaling these systems to handle real-time decision-making in dynamic environments without compromising performance was a significant barrier [25]. Additionally, the inherent uncertainties and variability in real-world settings—ranging from noisy sensors to unforeseen environmental changes—can complicate the learning process. Researchers are actively working on improving the efficiency and robustness of these methods to overcome these limitations and make them more suitable for complex, real-world applications.