2025-IEEE Access - Article 2

Safe Reinforcement Learning via Episodic Control

ZHUO LI, DERUI ZHU, JENS GROSSKLAGS

Abstract—Safe reinforcement learning (Safe RL) aims to learn policies capable of learning and adapting within complex environments while ensuring actions remain free from catastrophic consequences. This is a critical consideration in domains such as robotics, autonomous vehicles, and healthcare. Unlike traditional RL, which focuses mainly on maximizing episodic rewards, Safe RL integrates safety constraints to balance rewards and safety. Current Safe RL algorithms, while promising, lack sample efficiency as they require extensive environmental interactions to perform multi-objective optimization. This paper introduces an episodic-control-based method to enhance sample efficiency in safe policy optimization. In RL, methods based on episodic control involve the storing and replaying of past experiences or episodes, aiding in more efficient and precise policy optimization. Our proposed method involves clustering, measuring, and storing previous states based on a joint metric of returns and safety. Subsequently, we retrieve these state measurements and incorporate them into the policy optimization process through reward shaping. This approach effectively guides the policy towards high-return and safe decisions. We evaluate the performance of our method on established Safe RL benchmarks, including six safety-critical agent control tasks. The results demonstrate that our method can concurrently achieve higher episodic returns and fewer violations of safety constraints compared to the baseline methods, suggesting an effective balance between earning rewards and safety.

1. Contextual Challenges Introduction

Reinforcement Learning (RL) is effective in sequential decision-making but often risks unsafe actions. Safe RL addresses this by incorporating safety constraints. However, it faces **sample inefficiency** due to the need for extensive interactions to satisfy multiple objectives.

2. Related Works

In Safe Reinforcement Learning (Safe RL), previous research has focused on the following three approaches to improve sample efficiency:

2.1. Safety-Constrained Optimization

These methods explicitly integrate safety constraints into the optimization objectives, ensuring that safety boundaries are respected during learning.

2.2. Primal-Dual Methods

By transforming the original constraint problem into a dual problem, these methods address challenges where the original problem is difficult to optimize directly.

2.3. Model-Based Methods

By learning models of the environment, these approaches simulate potential safety issues to reduce sample requirements. While effective in reducing sample complexity, they often fail to fully utilize past interaction experiences.

Summary: The three aforementioned approaches have contributed to a reduction in sample demand to some extent, but they have not fully utilized past interaction experiences to enhance learning efficiency.

3. Proposed Work

This paper propose an episodic-control method to improve sample efficiency in Safe RL. This involves:

1.State abstraction to handle infinite state spaces.

2.Using reward and cost scores to evaluate states.

3.Joint optimization of these metrics to guide policy development.

4. My Perspective on This Paper

This paper addresses the issue of sample inefficiency in Safe Reinforcement Learning (Safe RL) by proposing an episodic-control-based Safe RL scheme. The approach focuses on leveraging past interaction experiences to enhance learning efficiency and innovatively constructs a closed-loop architecture of state abstraction → dual-dimensional measurement → memory-guided optimization.

This data-driven architectural design is highly instructive. The paper effectively improves sample utilization and reinforcement learning training speed by using an abstractor to abstract safety issues in the autonomous driving domain.